

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE MEDICINA**



**TESIS DOCTORAL**

**Evolución y biogeografía de los intrones de genes de  
trasplante en canarios silvestres**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR**

**PRESENTADA POR**

**Ester Muñoz Sánchez**

**Director**

**Antonio Arnaiz Villena**

**Madrid**

**Ed. electrónica 2019**

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE MEDICINA**

**DPTO. MICROBIOLOGÍA I**



**EVOLUCIÓN Y BIOGEOGRAFÍA DE LOS  
INTRONES DE GENES DE TRASPLANTE EN  
CANARIOS SILVESTRES**

**ESTER MUÑIZ SÁNCHEZ**

**DIRECTOR: ANTONIO ARNAIZ VILLENA**

**MADRID, 2019**



# **TESIS DOCTORAL**

## **EVOLUCIÓN Y BIOGEOGRAFÍA DE LOS INTRONES DE GENES DE TRASPLANTE EN CANARIOS SILVESTRES**

**AUTORA: ESTER MUÑIZ SÁNCHEZ**

**DIRECTOR: PROF. DR. ANTONIO ARNAIZ VILLENA**

Catedrático de Inmunología y Coordinador I+D Centro de Transfusión de la Comunidad de Madrid. Universidad Complutense de Madrid

### **LUGAR DE REALIZACIÓN:**

- Departamento de I+D del Centro de Transfusión de la Comunidad de Madrid
- Departamento de Microbiología I (Inmunología) de la Facultad de Medicina de la Universidad Complutense de Madrid



**Don Antonio Arnaiz Villena, Doctor en Medicina y Cirugía, Catedrático de Inmunología de la Universidad Complutense de Madrid y Coordinador de I+D del Centro de Transfusión de la Comunidad de Madrid,**

**CERTIFICA:**

**Que Doña Ester Muñoz Sánchez, ha realizado bajo su dirección el trabajo de Tesis Doctoral que lleva por título:**

**“Evolución y biogeografía de los intrones de genes de trasplante en canarios silvestres”.**

**Revisado el presente trabajo, considero que tiene la debida calidad para su defensa y calificación**

**Fdo.: Prof. Dr. Antonio Arnaiz Villena**



# **AGRADECIMIENTOS**

Me gustaría agradecer en primer lugar, al Dr. Antonio Arnaiz Villena, por haberme dado la oportunidad de trabajar en su equipo de investigación durante estos cuatro años y por haberme tutelado y dirigido en este camino que culmina con este proyecto de tesis.

A los compañeros que me han acompañado y asesorado durante todo mi proceso como doctorando, ayudándome a aprender las técnicas de laboratorio necesarias para el desarrollo de este proyecto de tesis.

Al Ministerio de Sanidad, por haber apoyado la investigación, sin el cual no se hubiera podido realizar este proyecto de tesis.

Al Centro de Transfusión de la Comunidad de Madrid, por haberme permitido el uso de sus instalaciones y a las personas que trabajan en él, por su amabilidad durante estos cuatro años.

Por último a la Universidad Complutense de Madrid, en especial al Departamento de Microbiología I de la Facultad de Medicina.





# DEDICATORIA

A Dani por estar siempre ahí cuando más te necesitaba, en los buenos y malos momentos. Por hacerme ver las cosas que realmente importan y ser el mejor compañero de vida que se puede tener.

A mis padres por creer en mí y apoyarme en todas las decisiones que he tomado en estos años.

A mi mejor amiga por toda su comprensión y apoyo. Por estar siempre ahí y haberte convertido en parte de mi familia.

Gracias a todos.

Os quiero.



# ÍNDICE



|  |           |
|--|-----------|
| <b>RESUMEN.....</b>  | <b>15</b> |
| <b>SUMMARY.....</b>  | <b>21</b> |
| <b>INTRODUCCIÓN.....</b>   | <b>27</b> |
| 1. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD.....  | 29        |
| 1.1 Definición y generalidades.....  | 29        |
| 2. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD EN HUMANOS (HLA).....                         | 31        |
| 2.1 Genes MHC.....   | 32        |
| 3. REGIÓN MHC DE CLASE I.....  | 35        |
| 3.1 Genes HLA de clase I clásicos.....   | 35        |
| 3.2 Genes HLA de clase I no clásicos.....  | 36        |
| 3.3 Estructura génica de los genes HLA de clase I.....                                       | 36        |
| 3.4 Moléculas HLA de clase I.....  | 37        |
| 3.5 Función de las moléculas HLA de clase I.....   | 40        |
| 4. REGIÓN MHC DE CLASE II.....   | 40        |
| 4.1 Genes HLA de clase II clásicos.....  | 40        |
| 4.2 Genes HLA de clase II no clásicos.....   | 41        |
| 4.3 Estructura génica de los genes HLA de clase II.....                                      | 41        |
| 4.4 Moléculas HLA de clase II.....   | 42        |
| 4.5 Función de las moléculas HLA de clase II.....  | 43        |
| 5. REGIÓN DE CLASE III.....  | 44        |
| 6. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD EN EL POLLO.....                              | 44        |
| 6.1 Historia y definición.....   | 44        |
| 6.2 Región MHC-Y.....  | 46        |
| 6.3 Región MHC-B.....  | 47        |
| 6.4 MHC y enfermedad en el pollo.....  | 52        |
| 7. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD EN OTRAS AVES.....                            | 54        |
| 7.1 Genes MHC en aves <i>Passeriformes</i> .....   | 55        |
| 7.2 MHC y enfermedad en aves <i>Passeriformes</i> .....                                      | 56        |
| 8. INTRONES.....   | 57        |
| 8.1 Historia y definición.....   | 57        |
| 8.2 Tipos de intrones.....   | 58        |
| 8.3 La función de los intrones.....  | 62        |
| 9. FILOGENIA MOLECULAR BASADA EN EL GEN DEL CIOCROMO B MITOCONDRIAL.....                     | 66        |
| 9.1 Modelos de sustitución nucleotídica.....   | 67        |
| 9.2 Métodos de construcción de árboles filogenéticos.....                                    | 68        |
| 10. LA CLASE AVES.....   | 70        |
| 10.1 Orden <i>Passeriformes</i> .....  | 71        |
| 10.2 El género <i>Serinus</i> .....  | 72        |
| 10.3 Aves incluidas en este estudio.....   | 72        |
| <b>OBJETIVOS.....</b>  | <b>85</b> |
| <b>MATERIALES Y MÉTODOS.....</b>   | <b>89</b> |
| 1. MUESTRAS EMPLEADAS.....   | 91        |
| 1.1 Muestras para el estudio del intrón 2 de los genes MHC de clase I.....                   | 91        |
| 1.2 Muestras para el estudio filogenético basado en el gen del citocromo b mitocondrial..... | 92        |
| 2. OBTENCIÓN DE MATERIAL GENÉTICO.....   | 93        |

|   |     |
|---|-----|
| 2.1 Extracción de ADN.....  | 94  |
| 3. OBTENCIÓN DE SECUENCIAS DE ADN DEL INTRÓN 2 DE LOS GENES MHC DE CLASE I.....                     | 96  |
| 3.1 Amplificación del intrón 2 de los genes MHC de clase I.....                                     | 96  |
| 3.2 Control de la amplificación mediante electroforesis en gel de agarosa.....                      | 99  |
| 3.3 Purificación de los productos de PCR.....   | 99  |
| 3.4 Secuenciación de los productos purificados.....   | 100 |
| 3.5 Clonación de los productos purificados.....   | 101 |
| 3.6 Obtención de las secuencias de ADN de los fragmentos clonados.....                              | 103 |
| 4. ANÁLISIS DE LAS SECUENCIAS DE ADN DEL INTRÓN 2 DE LOS GENES MHC DE CLASE I.....                  | 104 |
| 4.1 Caracterización de las secuencias de estudio.....   | 104 |
| 4.2 Análisis del grado de conservación de las secuencias de estudio.....                            | 105 |
| 5. FILOGENIA DEL GÉNERO <i>SERINUS</i> BASADA EN EL GEN DEL CITOCROMO B MITOCONDRIAL.....           | 105 |
| 5.1 Construcción de un árbol filogenético basado en el método <i>Neighbor-Joining</i> (NJ).....     | 105 |
| 5.2 Construcción de un árbol filogenético basado en el método <i>Maximum Likelihood</i> (ML).....   | 106 |
| 5.3 Construcción de un árbol filogenético basado en el método Bayesiano (BI).....                   | 106 |
| <b>RESULTADOS</b> .....   | 107 |
| 1. CARACTERIZACIÓN DE LAS SECUENCIAS DE ADN DEL INTRÓN 2 DE LOS GENES MHC DE CLASE I.....           | 111 |
| 1.1 Frecuencias nucleotídicas de las secuencias del intrón 2.....                                   | 117 |
| 1.2 Distancias genéticas entre las secuencias del intrón 2 .....                                    | 118 |
| 1.3 Porcentaje de similitud entre las secuencias del intrón 2.....                                  | 123 |
| 2. FILOGENIA MOLECULAR DEL GÉNERO <i>SERINUS</i> BASADA EN EL GEN DEL CITOCROMO B MITOCONDRIAL..... | 125 |
| 2.1 Árbol filogenético basado en el método <i>Neighbor-Joining</i> (NJ).....                        | 126 |
| 2.2 Árbol filogenético basado en el método <i>Maximum Likelihood</i> (ML).....                      | 127 |
| 2.3 Árbol filogenético basado en el método Bayesiano (BI).....                                      | 128 |
| <b>DISCUSIÓN</b> .....  | 131 |
| 1. FILOGENIA MOLECULAR Y EVOLUCIÓN DEL GÉNERO <i>SERINUS</i> .....                                  | 133 |
| 2. EL INTRÓN 2 DE LOS GENES MHC DE CLASE I EN LAS AVES SILVESTRES DEL GÉNERO <i>SERINUS</i> .....   | 137 |
| 2.1 Análisis de las secuencias del intrón 2 obtenidas.....  | 138 |
| 2.2 Análisis del porcentaje de similitud entre las secuencias del intrón 2.....                     | 142 |
| 2.3 Análisis de las frecuencias nucleotídicas de las secuencias del intrón 2.....                   | 145 |
| 2.4 La importancia del estudio de los intrones.....   | 146 |
| <b>CONCLUSIONES</b> .....   | 149 |
| <b>BIBLIOGRAFÍA</b> .....   | 153 |
| <b>ANEXOS</b> .....   | 169 |
| Anexo 1: Abreviaturas más frecuentes.....   | 171 |
| Anexo 2: Curriculum Vitae.....  | 172 |

# RESUMEN





### **Evolución y biogeografía de los intrones de genes de trasplante en canarios silvestres**

El Sistema Principal de Histocompatibilidad (en inglés MHC de *Major Histocompatibility Complex*) es el sistema genético más polimórfico existente en los animales. Consiste principalmente en varios loci que codifican glicoproteínas de superficie, cuya función más conocida es la de activar la maquinaria del sistema inmune mediante la presentación antigénica. En los seres humanos este sistema se denomina Sistema HLA (del inglés *Human Leukocyte Antigens*). Dentro de este sistema destacan las moléculas MHC de clase I y de clase II. A lo largo de los años los genes MHC se han ido secuenciando en especies de mamíferos y en algunas especies de aves, así como en otros organismos, valiéndose de los avances en las técnicas de secuenciación del ADN. En el caso de las aves, el MHC del pollo (*Gallus gallus*) ha sido uno de los más estudiados, debido a su interés para la industria alimentaria. Su MHC fue catalogado como “MHC mínimo esencial”, debido a que este sistema es más pequeño comparado con el de los mamíferos y a que tiene intrones muy cortos, entre otras razones. Además esta suposición se extendió al resto de especies de la clase Aves. Sin embargo, en la actualidad y gracias a nuevas investigaciones, este concepto de “MHC mínimo esencial” ya no se suele aplicar al MHC de todas las aves estudiadas ya que poco a poco se van describiendo nuevos genes MHC tanto de clase I como de clase II que no siguen las características descritas para el pollo. El estudio de los intrones, ha generado bastante interés en los últimos años, ya que se ha visto que no se trata de “ADN basura”, sino que esos fragmentos de ADN juegan un papel muy importante en la expresión de muchos genes. El estudio de los intrones en aves silvestres de campo y su comparación con otras aves domésticas como es el caso del pollo, proporciona una visión más natural de la evolución de estos fragmentos a lo largo de millones de años. Los objetivos que se plantean para este proyecto de tesis son: 1) Obtener las secuencias de ADN del intrón 2 de los genes MHC de clase I entre las aves silvestres de canto de un mismo género: el género *Serinus*. 2) Estudiar la evolución de la conservación del intrón 2 de los genes MHC de clase I mediante el análisis de las secuencias de ADN para determinar si hay posiciones variables, su localización y si estas posiciones se dan también en la secuencia del intrón 2 de otras aves silvestres y domésticas. 3) Estudiar la composición nucleotídica de la secuencia del intrón 2 de los genes MHC de clase I y compararla con la de otras aves. 4) Determinar si las aves de canto del género *Serinus* poseen un “MHC mínimo esencial” que implica que los intrones son inusualmente cortos, como es el caso

del pollo. 5) Realizar un análisis evolutivo filogenético de especies hermanas basado en secuencias del gen del citocromo b mitocondrial, para ver su correlación con la evolución del intrón 2 de los genes MHC de clase I y comprobar si siguen un patrón de evolución similar.

En este proyecto de tesis se han obtenido datos moleculares procedentes de aves silvestres pertenecientes al género *Serinus*. Se ha secuenciado la región correspondiente al intrón 2 de los genes MHC de clase I de 11 especies pertenecientes a este género. Además se han utilizado las secuencias del citocromo b mitocondrial de 19 especies del género *Serinus* para estudiar su evolución en el tiempo y poder así establecer relaciones entre las diferentes especies que forman este género.

Los resultados de este proyecto de tesis muestran que la secuencia del intrón 2 de los genes MHC de clase I de las aves pertenecientes al género *Serinus* está altamente conservada, mostrando sólo 28 nucleótidos diferentes de un total de 309. El canario silvestre (*Serinus canaria*) es la excepción del grupo de canarios, ya que posee 183 nucleótidos diferentes con respecto al resto del grupo de su mismo género, además de que su secuencia es algo más corta que la del resto, contando sólo con 286 nucleótidos. La secuencia más diferente de todo el análisis es la del pollo (*Gallus gallus*), la cual de un total de 229 nucleótidos, 173 son diferentes al resto del grupo. Estas diferencias también se aprecian en los análisis de las distancias genéticas y el porcentaje de similitud.

Las frecuencias nucleotídicas de las secuencias del intrón 2 de los genes MHC de clase I de las aves pertenecientes al género *Serinus* son muy similares entre sí, al igual que el pinzón vulgar (*Fringilla coelebs*). Sin embargo el pollo (*Gallus gallus*) destaca por sus porcentajes de Timina, Citosina y Adenina. El porcentaje de Timina del pollo es de 11,8%, mientras que el del resto del grupo está alrededor del 25%. En el caso del porcentaje de Citosina, el pollo tiene un 42,4% de la misma, mientras que el resto de las aves sólo tienen un 18% aproximadamente. Por último destacar el bajo porcentaje de Adenina que posee el pollo (9,6%), con respecto al 38% del resto de especies. Debido a estos valores el porcentaje de G+C del pollo (78,6%) también varía con respecto al resto de especies del análisis (55%).

En lo que respecta a la evolución del género *Serinus*, los resultados muestran que la edad aproximada de aparición de estas aves en la tierra es de unos 9 millones de años. Posteriormente, entre 1 y 2 millones de años después se dividieron en dos grupos principalmente. Uno de ellos dio lugar a los canarios africanos, que a su vez se dividió

en tres grupos, los canarios africanos basales (*S. striolatus*, *S. totta* y *S. thibetanus*), que serían los más antiguos; los canarios grandes africanos (*S. sulphuratus*, *S. gularis*, *S. flaviventris* y *S. albogularis*); los canarios pequeños africanos (*S. leucopygius*, *S. citrinelloides*, *S. dorsostratus*, *S. atrogularis*, *S. citrinipectus* y *S. mozambicus*). El otro grupo dio lugar a los canarios que abandonaron África para colonizar otras regiones de la tierra. En este grupo se incluyen los canarios afroasiáticos (*S. canicollis*, *S. alario*, *S. syriacus* y *S. pusillus*) y los canarios mediterráneos (*S. serinus* y *S. canaria*).

Las conclusiones de este proyecto de tesis serían: 1) El tamaño de la secuencia del intrón 2 de los genes MHC de clase I de las aves de canto pertenecientes al género *Serinus* es mayor que el de las secuencias del intrón 2 del pollo (*Gallus gallus*). Por tanto, el tamaño del intrón 2 de nuestro estudio que cuenta con 80 nucleótidos más que el intrón 2 del pollo, no permite establecer que tengan un MHC “mínimo esencial”, ni que este concepto sea universal en las aves. 2) Existe una conservación de la secuencia del intrón 2 de los genes MHC de clase I de las aves de canto del género *Serinus* de al menos 9 millones de años, lo que sugiere que esta secuencia juega un papel importante en la expresión de los genes de Histocompatibilidad y la existencia de una presión evolutiva positiva indeterminada sobre este intrón. 3) El intrón 2 de los genes MHC de clase I ha evolucionado de forma diferente al gen del citocromo b mitocondrial, como reflejan los análisis filogenéticos realizados con estos dos tipos de secuencias, como era de esperar por el funcionalismo conocido de ambas moléculas. 4) Mediante el análisis filogenético de las secuencias del citocromo b mitocondrial se ha realizado un estudio de la evolución de dicho gen en las aves de canto del género *Serinus*, lo que da una idea de sus tiempos de parentesco genético y especiación a lo largo de millones de años desde su aparición en la Tierra.



# SUMMARY



The Major Histocompatibility Complex is the most polymorphic genetic system existing in animals. It consists mainly of several loci that encode surface glycoproteins, whose main function is to activate the machinery of the immune system through antigen presentation. In humans, this system is called HLA System. Within this system are MHC class I and class II molecules. Over the years, MHC genes have been sequenced in mammalian species and in some bird species, as well as in other organisms, using advances in DNA sequencing techniques. In the case of birds, the chicken MHC has been one of the most studied, due to its interest for the food industry. His MHC was cataloged as "minimum essential MHC", because this system is smaller compared to that of mammals and has very short introns, among other features. In addition, this assumption was extended to the rest of the species of the Aves class. However, at present and due to new investigations, this concept of "minimum essential MHC" is no longer applied to the MHC of all the birds studied as new MHC genes of both class I and class are gradually being described. The study of intron sequences has generated much interest in recent years, since it has been seen that it is not "junk DNA", but these fragments of DNA play a very important role in the expression of many genes. The study of introns in wild birds and their comparison with other domestic birds such as chicken provides a more natural view of the evolution of these fragments over millions of years.

The aims of this thesis project are: 1) To compare intron 2 DNA sequence among genus *Serinus* wild birds; 2) To study the conservation level of intron 2 by analyzing DNA sequences to determine if there are variable positions, their location and if these positions are also given in the sequence of intron 2 of other birds ; 3) To study the nucleotide composition of the intron 2 sequence and compare it with that of other birds; 4) To determine if birds of the genus *Serinus* have an "minimum essential MHC" in terms of the factor of intron length, as is the case of chicken; 5) To perform a phylogenetic analysis based on mitochondrial cytochrome b gene sequences, to see its correlation with the evolution of intron 2 and check if they follow a similar evolution pattern.

In this thesis project we have obtained molecular data from wild birds belonging to genus *Serinus*. The region corresponding to the intron 2 of the MHC class I genes of 11 species belonging to this genus has been sequenced. In addition, the mitochondrial cytochrome b sequences of 19 species of the genus *Serinus* have been used to study



their evolution over time and thus be able to establish relationships between the different species from this genus.

The results of this thesis project show that the sequence of intron 2 of birds belonging to the genus *Serinus* is highly conserved, showing only 28 different nucleotides out of a total of 309. The wild canary (*Serinus canaria*) is the exception of the group of canaries, since it has 183 different nucleotides with respect to the rest of the group of the same genus, and its sequence is shorter than the rest, counting only with 286 nucleotides. The most different sequence of the whole analysis is the chicken's one, which out of a total of 229 nucleotides, 173 are different from the rest of the group. These differences are also seen in the analysis of genetic distances and the percentage of similarity.

The nucleotide frequencies of the intron 2 sequences of the birds belonging to the genus *Serinus* are very similar to each other, as the common chaffinch (*Fringilla coelebs*). However, the chicken (*Gallus gallus*) stands out for its percentages of thymine, cytosine and adenine. The percentage of thymine is 11.8%, while that of the rest of the group is around 25%. In the case of the percentage of cytosine, the chicken has 42.4% of it, while the rest of the birds only have about 18%. Finally, we should point out the low percentage of adenine (9.6%), compared to 38% in the rest of the species. Due to these values the percentage of (G + C) of the chicken (78.6%) also varies with respect to the rest of the species of the analysis (55%).

Regarding to the evolution of the genus *Serinus*, the results show that the approximate age of appearance of these birds on earth is about 9 million years. Subsequently, between 1 and 2 million years later they were divided mainly into two groups. One of them gave rise to the African canaries, which in turn was divided into three groups, the basal African canaries (*S. striolatus*, *S. totta* and *S. thibetanus*), which would be the oldest; the large African canaries (*S. sulphuratus*, *S. gularis*, *S. flaviventris* and *S. albogularis*); and the small African canaries (*S. leucopygius*, *S. citrinelloides*, *S. dorsostriatus*, *S. atrogularis*, *S. citrinipectus* and *S. mozambicus*). The other group gave rise to canaries that left Africa to colonize other regions. This group includes the Afro-Asian canaries (*S. canicollis*, *S. alario*, *S. syriacus* and *S. pusillus*) and the Mediterranean canaries (*S. serinus* and *S. canaria*).

The conclusions of this thesis project would be: 1) The size of the intron 2 sequence of the MHC class I genes of the songbirds belonging to the genus *Serinus* is greater than the sequences of the intron 2 of the chicken (*Gallus gallus*), so regarding to the size of the intron can not be established that these birds have an "minimum essential MHC" 2) There is a conservation of the sequence of intron 2 of MHC class I genes of genus *Serinus* birds of at least 9 million years, which suggests that this sequence plays an important role in the expression of histocompatibility genes and the existence of a positive evolutionary pressure over this intron. 3) Intron 2 of the MHC class I genes has evolved differently from the mitochondrial cytochrome b gene, as reflected in the phylogenetic analyzes performed with these two types of sequences. 4) Through the phylogenetic analysis of mitochondrial cytochrome b sequences, a study of the evolution of this gene in birds from genus *Serinus* has been carried out, which gives an idea of its evolution and speciation over millions of years.



# INTRODUCCIÓN



## 1. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD

### 1.1 Definición y generalidades

El Sistema Principal de Histocompatibilidad (SPH, en inglés *MHC Major Histocompatibility Complex*) es un sistema de genes implicado en la activación de la respuesta inmune adaptativa. Es altamente polimórfico y está presente en todos los vertebrados (Klein 1986; Humphreys & Reinherz 1994). Este sistema va a codificar para una serie de glicoproteínas de superficie cuya función es presentar antígenos (péptidos derivados de agentes infecciosos, como pueden ser los péptidos microbianos) a las células efectoras del sistema inmune, como los linfocitos T, las cuales una vez activadas, son capaces de poner en marcha una respuesta inmune específica contra el agente infeccioso causante de su activación (Zinkernagel & Doherty 1979). En humanos se denomina Sistema HLA (del inglés, *Human Leukocyte Antigen*).

Los genes de histocompatibilidad fueron identificados por primera vez en la década de los 40, gracias a la unión de dos campos de investigación inicialmente separados: uno dedicado al estudio del cáncer, mediante el trasplante de tumores en ratones y otro relacionado con la identificación de grupos sanguíneos. En los inicios del siglo XX se había constatado que los tumores trasplantados en ratones pertenecientes a cepas diferentes eran rechazados por el receptor (Jensen 1903; Loeb 1908) ocurriendo lo mismo cuando era tejido sano el que se trasplantaba (Little & Johnson 1922). En la década de los 30, J.B. Haldane sugirió que la destrucción de los tejidos trasplantados por parte del receptor podría tratarse de un proceso similar al de la lisis eritrocitaria y al de la hemaglutinación, cuando se ponen en contacto eritrocitos de un individuo con suero de otro individuo previamente inmunizado con esos eritrocitos, en el caso de la lisis, y sin inmunizar, en el caso de la hemaglutinación. Años más tarde se identificaron varios antígenos eritrocitarios en ratones receptores de trasplante, relacionándose uno de ellos con la supervivencia del tumor trasplantado, al que se denominó como antígeno II. Este hecho permitió elaborar una teoría inmunológica del trasplante (Gorer 1936). Una década más tarde, G. Snell identificó la región génica responsable de este rechazo, denominando a los genes que la componían, genes de histocompatibilidad (Snell 1948). En el caso concreto del ratón se denominaron genes H-2, en referencia al antígeno II.

Los genes de histocompatibilidad en humanos fueron descubiertos por J. Dausset cuando detectó la presencia de anticuerpos capaces de provocar la aglutinación de los leucocitos del donante pero no los del receptor, al reconocer los antígenos en la

superficie de estas células, en personas que habían recibido múltiples transfusiones sanguíneas (Dausset 1958). Gracias a este descubrimiento, estas moléculas se denominaron con las siglas HLA (*Human Leukocyte Antigen*). En la década de los 60 se comprobó que estos antígenos leucocitarios representaban el Sistema Principal de Histocompatibilidad (Dausset et al. 1965; Ceppellini et al. 1967). Actualmente se conoce que la función principal de las moléculas HLA es la presentación antigénica mediante su interacción con los linfocitos T (Stern et al. 1994; Madden 1995). Este papel fundamental de las moléculas HLA involucradas en la respuesta inmune frente a antígenos fue propuesto en 1970 cuando se demostró que los linfocitos T no eran capaces de reconocer antígenos libres en la circulación, sino que necesitaban de estas moléculas para poder reconocerlos y procesarlos.

En la actualidad se puede decir del MHC que es un sistema cuya función principal es reconocer específicamente péptidos extraños (y no propios) y presentarlos al sistema inmune para que actúe y elimine el origen de dichos péptidos, ya sea un agente infeccioso o una célula tumoral (Klein 1986). Es indudable la implicación sanitaria de este sistema en cuanto a su relación con las enfermedades infecciosas y con los tumores, pero no es menor su relevancia, en el caso humano, en los trasplantes de órganos y en su estrecha vinculación con enfermedades de tipo autoinmune. A pesar de su función inmunológica, y por motivos históricos, el Sistema Principal de Histocompatibilidad ha conservado este nombre por su relación con la compatibilidad entre tejidos.

El Sistema Principal de Histocompatibilidad presenta 2 características fundamentales que dificultan la evasión por parte de los patógenos del sistema inmune del huésped.

**Es poligénico:** Existen varios genes que codifican para moléculas del MHC de clase I y de clase II, codificando cada uno de ellos proteínas con un rango de especificidad peptídico diferente.

**Es muy polimórfico:** Existen multitud de alelos para cada gen del sistema. De hecho el MHC es el sistema más polimórfico conocido.

Otras características de este sistema son:

**Posee desequilibrio de ligamiento:** Fenómeno por el cual diferentes alelos de distintos genes se encuentran en el mismo cromosoma con la frecuencia mayor a la teóricamente esperada en una combinación al azar. Esto tiene como consecuencia que se imposibilite la transferencia de ciertos genes a la progenie de manera aleatoria. Hay zonas en la región de clase II, en las que no se ha observado nunca recombinación a pesar del elevado número de familias que se ha estudiado. Este es el caso del tramo presente entre DRA y DQA, que se hereda en bloque. A pesar de que existen zonas del MHC en las que nunca se produce recombinación, parece ser que hay otras en las que este fenómeno es bastante frecuente. Una de estas zonas, también denominada “punto caliente” es la zona de recombinación entre DP y DQ. También se cree que existen más “puntos calientes” en la regiones de clase I y clase III (Thomsen et al. 1994).

**Su expresión es codominante:** Se expresan los alelos de procedencia paterna y los de procedencia materna.

## 2. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD EN HUMANOS (HLA)

El sistema HLA se localiza en el brazo corto del cromosoma 6, en la parte distal de la banda 6p21.1-21.3. Se considera una de las regiones con mayor densidad génica dentro del genoma humano, ya que presenta un elevado número de loci distribuidos en una extensión de casi 4 centimorgan, aproximadamente 4.000.000 de pares de bases (pb) (Berger et al. 1979).

El primer mapa genético de esta región genómica se completó en 1991 (Trowsdale et al. 1991). Años más tarde se elaboró otro mapa más completo a partir de secuencias parciales en individuos diferentes, según el cual la región HLA ocupa 3.6 Mb y está compuesta por 224 genes, de los cuales 128 podrían expresarse y aproximadamente el 40% tendrían función inmunológica, que se reparten en tres regiones bien definidas: región HLA de clase I, región HLA de clase II y región HLA de clase III (The MHC sequencing consortium 1999). Este mapa fue posteriormente actualizado por una secuencia completa obtenida a partir de una línea celular consanguínea y homocigótica para todos los genes del sistema (Stewart et al. 2004).

La secuenciación completa del cromosoma 6 (Mungall et al. 2003) reveló la existencia de otros genes relacionados con MHC más allá de las 3.6 Mb descritas en 1999, por lo que se elaboró un nuevo mapa genético que recibió el nombre de MHC



extendido (xMHC, del inglés *extended MHC*) (Horton et al. 2004) para diferenciarlo del mapa de 1999/2004, que actualmente se denomina mapa clásico (The MHC sequencing consortium 1999).

El “*MHC Haplotype Project*” ha confirmado la existencia de más de 44.000 variaciones, tanto sustituciones como indels (inserciones y deleciones) y la presencia de más de 300 loci en este sistema, de los cuales al menos 160 codifican para proteínas. Dentro de este grupo, 122 genes codificarían sustituciones, de las cuales 97 son no sinónimas (Horton et al. 2008).

El mapa clásico está dividido en 3 regiones denominadas HLA de clase I, HLA de clase III y HLA de clase II, situadas en este orden de telómero a centrómero en el brazo corto del cromosoma 6. El mapa extendido completa el anterior al añadir dos nuevas regiones: una en el extremo telomérico (región de clase I extendida) y otra en el extremo centromérico (región de clase II extendida), ampliándose así las dimensiones en la región HLA hasta las 7.6 Mb. Según este mapa extendido, la región HLA se sitúa entre las posiciones cromosómicas 6p21.1-6p21.3 y contiene 421 genes, de los cuales el 60% expresan proteínas funcionales, el 33% son pseudogenes (expresan proteínas aberrantes o bien no se expresan) y el 7% restante se transcriben dando lugar a moléculas de ARN transferente (ARNt), que participan en el proceso de traducción del ARN mensajero (ARNm) en el ribosoma. La mayor concentración de genes se observa en la región de clase III, que contiene 58 genes funcionales (23%) en 0.7 Mb, lo que la convierte además en la región de mayor densidad génica de todo el genoma humano (Horton et al. 2004), mientras que la región de clase I extendida recoge el mayor número de genes funcionales (109, 43% del total de estos genes) del sistema.

### 2.1 Genes MHC

Como se ha comentado en apartados anteriores, el Sistema Principal de Histocompatibilidad está formado por una serie de genes que se distribuyen en varias regiones. Estas regiones son: clase I extendida, clase I, clase III, clase II y la clase II extendida. Dentro de estas regiones existen varios grupos de genes, los cuales serán comentados a continuación (Horton et al. 2004):

**Histonas:** estos genes codifican proteínas involucradas en la formación del núcleo proteico del nucleosoma, alrededor del cual se pliega el ADN para conformar la cromatina. La mayor concentración de estos genes en el genoma humano se encuentra en la región HLA de clase I extendida, que contiene 55 genes que expresan histonas de

las 5 clases existentes (6 H1, 12 H2A, 15 H2B, 10 H3 y 12 H4) y 11 pseudogenes distintos.

**ARN transferente (ARNt):** son genes muy cortos (75-90 nucleótidos) y codifican moléculas de ARN implicadas en la síntesis proteica, ya que reconocen específicamente los tripletes del ARNm y aportan el aminoácido correspondiente. Son 157 genes (6 de ellos no funcionales) que incluyen todo los tipos existentes salvo Asn- y Cys-ARNt. Están distribuidos a lo largo de la región de clase I extendida y representan el mayor agrupamiento de este tipo de genes de todo el genoma humano.

**Transportadores de solutos:** estos genes aparecen en las regiones de clase I y II extendidas y codifican proteínas transportadoras de nutrientes cuyo papel fisiológico es aún desconocido.

**Butirofilinas:** son miembros de la superfamilia de las inmunoglobulinas, presentes en la región de clase I extendida. Se trata de 8 genes cuya función es desconocida, aunque se cree que están involucrados en el metabolismo de los lípidos (BTN2A2) y en la secreción de estos en la leche (BTN1A1). Son muy similares a los genes MOG de los mamíferos (situado entre las regiones de clase I y clase III en humanos) y a los genes B-G del sistema principal de histocompatibilidad del pollo.

**Receptores vomeronasales:** son genes que se localizan en la región de clase I extendida y pertenecen al grupo de genes que codifica para receptores de feromonas.

**Receptores olfatorios:** se trata de un conjunto de 34 genes, 14 de ellos funcionales, que se encuentran en la región de clase I extendida.

**Dedos de zinc:** estos genes codifican proteínas muy diversas que se agrupan por la presencia de dominios capaces de unir iones de zinc. Los productos de estos genes presentan funciones muy diversas, pudiendo actuar como enzimas, factores de transcripción y proteínas de replicación. En la región de clase I se localizan 36 de estos genes, los cuales se dividen a su vez en tres grupos: los que tienen un dominio SCAN (13 genes), los que tienen un dominio triple denominado TRIM (8 genes) y el resto (15 genes).

**HLA-I:** son genes que se caracterizan por expresar proteínas estructuralmente muy similares, aunque diferentes desde el punto de vista funcional. Según esto, se clasifican en genes de clase I clásicos (HLA-A, HLA-B y HLA-C, también denominados HLA-Ia), genes de clase I no clásicos (HLA-E, HLA-F, HLA-G y 12 pseudogenes también conocidos como HLA-Ib), genes tipo HLA-I (MIC-A, MIC-B y 5 pseudogenes) y el gen de la hemocromatosis (HFE). Todos están localizados en la

región de clase I, excepto el gen HFE que se encuentra en el extremo telomérico de la región de clase I extendida.

**Factores de necrosis tumoral:** estos genes codifican 3 tipos de citoquinas (TNF, LTA y LTB) pertenecientes a la superfamilia TNF y que participan en procesos inflamatorios. Se encuentran situados en la región de clase III.

**Antígeno linfocitario 6:** son genes que codifican proteínas de membrana asociadas a glicosil-fosfatidil-inositol y pueden tener una función inmunológica. Están localizados en la región de clase III.

**Proteínas de choque térmico:** los genes que codifican para este tipo de proteínas aumentan su expresión en situaciones de estrés celular, como un choque térmico, colaborando en los procesos de síntesis, plegamiento, ensamblaje, transporte y degradación de proteínas. Concretamente este grupo que se encuentra en la región de clase III está involucrado en la señalización inducida por estrés del sistema inmune para la eliminación de células dañadas, infectadas o tumorales.

**HLA-II:** son genes que se localizan en la región de clase II y dan lugar a proteínas polimórficas implicadas en la presentación antigénica que se denominan HLA-DM, HLA-DO (conocidos también como genes de clase II no clásicos), HLA-DP, HLA-DQ y HLA-DR (conocidos como genes de clase II clásicos). Los genes clásicos se expresan en células presentadoras de antígenos (APC, del inglés *Antigen-Presenting Cells*), fundamentalmente células dendríticas (DC, del inglés *Dendritic Cells*), macrófagos y linfocitos B. Su función es unir péptidos de origen extracelular (generalmente procedentes de bacterias) y presentarlos a linfocitos T cooperadores (Th, CD4+) para poner en marcha una respuesta inmunológica de tipo humoral, mediada por anticuerpos. Los genes no clásicos no se expresan en la superficie celular, pero forman complejos que están involucrados en el intercambio y carga de péptidos en las moléculas de clase II clásicas.

Algunos genes de complemento de la vía clásica (C2, C4A, C4B) y de la vía alternativa (Bf) también se pueden encontrar en la región de clase III.

Una característica fundamental de la región HLA es que los genes que contiene están muy próximos entre sí, generando desequilibrios de ligamiento muy acentuados (la recombinación entre ellos es altamente improbable) que provocan la existencia de haplotipos (combinaciones de alelos concretos situados en loci contiguos) que se transmiten como unidades genéticas a lo largo de muchas generaciones.

### 3. REGIÓN MHC DE CLASE I

Esta región del MHC está situada en la parte más telomérica del MHC en el cromosoma 6, comprendiendo aproximadamente 1.9 Mb. En ella se localizan los genes HLA de clase I (Figura I1) (Horton et al. 2004).

Normalmente cuando se habla de los genes de clase I se está haciendo alusión a las proteínas que resultan de la expresión de los genes HLA de clase I clásicos (HLA-A, -B, -C) y no clásicos (HLA-E, -F, -G), quedando excluidas las proteínas expresadas por los genes MIC y el gen HFE.

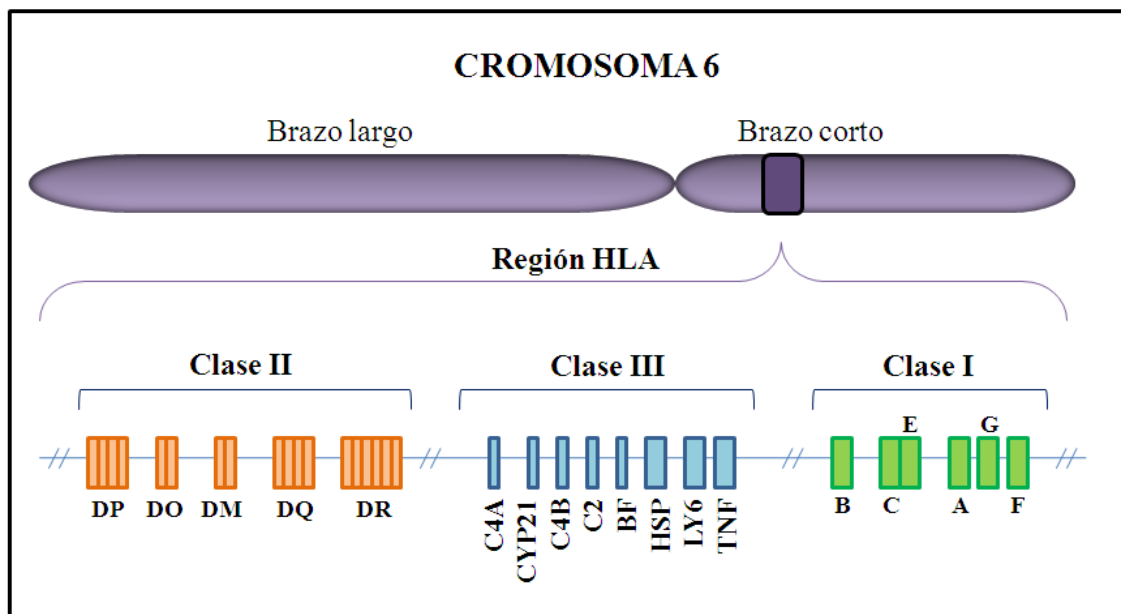


Figura I1: Mapa simplificado del MHC donde se resaltan las regiones de clase I, clase II y clase III.

#### 3.1 Genes HLA de clase I clásicos

Dentro de este grupo se encuentran los genes HLA-A, HLA-B y HLA-C. Codifican para glicoproteínas de membrana que se expresan en todas las células nucleadas del organismo. Su nivel de expresión varía desde un máximo en células pertenecientes al sistema inmune, como son linfocitos B y T y macrófagos, hasta un mínimo en células musculares, fibroblastos y células del sistema nervioso.

Las proteínas codificadas por los genes HLA clásicos de clase I están implicadas en el reconocimiento de péptidos antigénicos de origen intracelular, que son presentados a los linfocitos T citotóxicos (Tc, CD8+) para poner en marcha una respuesta

inmunológica de tipo celular mediada por linfocitos Tc y células NK (del inglés, *Natural Killer*).

### **3.2 Genes HLA de clase I no clásicos**

Dentro de este grupo se incluyen los genes HLA-E, -F y -G. Codifican para proteínas estructuralmente similares a las de los genes clásicos anteriormente comentados. A diferencia de los genes HLA de clase I clásicos, tienen una expresión tisular más limitada. Todas se expresan en las células del trofoblasto, además HLA-G se expresa también en las células epiteliales del timo en adultos, en la matriz de las uñas y en la córnea, mientras que HLA-F se expresa en células B y en los tejidos linfoides (Gomez-Prieto et al. 2010).

Por otro lado tienen un polimorfismo bajo, ya que el locus más polimórfico de este grupo, HLA-G, presenta sólo 61 alelos distintos descritos hasta la fecha. Este número contrasta mucho con los 3930 alelos descritos para HLA-C, el locus menos polimórfico de los genes HLA de clase I clásicos (Robinson et al. 2015).

Se teoriza sobre su función tolerogénica, al tratarse de moléculas inmunosupresoras a varios niveles de la respuesta inmune. Además teniendo en cuenta su expresión tisular, tendrían un papel clave a la hora de evitar el rechazo del feto durante el embarazo (Gomez-Prieto et al. 2010).

### **3.3 Estructura génica de los genes HLA de clase I**

Los genes HLA de clase I están constituidos por 8 exones y 7 intrones (Malissen et al. 1982; Jordan et al. 1985). Los exones codifican para diferentes dominios o regiones de la molécula HLA.

**El exón 1** codifica para la región 5'UT (del inglés *Untranslated*) y para un péptido líder de 24 aminoácidos que contiene un núcleo hidrófobo característico de las proteínas que migran desde el retículo endoplasmático. **Los exones 2 y 3** son los más polimórficos y codifican respectivamente para los dominios  $\alpha 1$  y  $\alpha 2$  de la molécula a los que se une el péptido antigénico. En los genes HLA de clase I clásicos, el polimorfismo de estos exones es el responsable de la especificidad en la unión de los péptidos antigénicos a la molécula HLA. **El exón 4** es menos polimórfico y codifica para el dominio  $\alpha 3$ . **El exón 5** codifica para el dominio transmembrana hidrofóbico.

El dominio citoplasmático de los genes HLA de clase I clásicos está codificado por los **exones 6 y 7** y una parte del **exón 8** (en las moléculas HLA-A y HLA-C). El exón 8 codifica además para la región 3'UT. En el caso de HLA-E, este dominio está codificado únicamente por los exones 6 y 7, además presenta una delección de 5

nucleótidos originándose un codón de terminación, por lo que este dominio es más corto en comparación con el de HLA-B. En HLA-F el dominio citoplasmático está codificado por los exones 6 y 8, ya que el exón 7 se elimina mediante *splicing* (mecanismo de corte y empalme). Por último tenemos a HLA-G, el cual posee un dominio citoplasmático muy corto, de sólo 6 aminoácidos, 5 de ellos codificados por el exón 5 y uno codificado por el exón 6, ya que en el segundo codón del exón 6 es un codón de terminación.

### ***3.4 Moléculas HLA de clase I***

Las moléculas HLA de clase I, son proteínas que resultan de la expresión de los genes HLA de clase I clásicos (HLA-A, -B y -C) y no clásicos (HLA-E, -F y -G). Se trata de glicoproteínas de superficie pertenecientes a la superfamilia de las inmunoglobulinas, compuestas por dos cadenas polipeptídicas: una cadena pesada o cadena alfa ( $\alpha$ ) de 44 kilodalton (Kd) codificada en la región de clase I del sistema HLA (brazo corto del cromosoma 6) y una cadena ligera o cadena beta ( $\beta$ ) de 12 kilodalton (Kd), denominada beta-2-microglobulina ( $\beta$ 2M) (Figura I2) (Grey et al. 1973; Peterson et al. 1974) codificada fuera del sistema HLA, en el cromosoma 15 (Ploegh et al. 1981). Ambas cadenas se unen no covalentemente en su porción extracelular, siendo la cadena  $\alpha$  es la única que atraviesa la membrana celular.

La porción extracelular de la cadena pesada, o cadena  $\alpha$ , se divide en tres dominios, denominados dominios  $\alpha$ 1,  $\alpha$ 2 y  $\alpha$ 3, codificados por los exones 2, 3 y 4 respectivamente (Malissen et al. 1982). La porción transmembrana de esta cadena, forma una estructura en  $\alpha$ -hélice que continúa con un pequeño tallo citoplasmático rico en tirosinas y serinas.

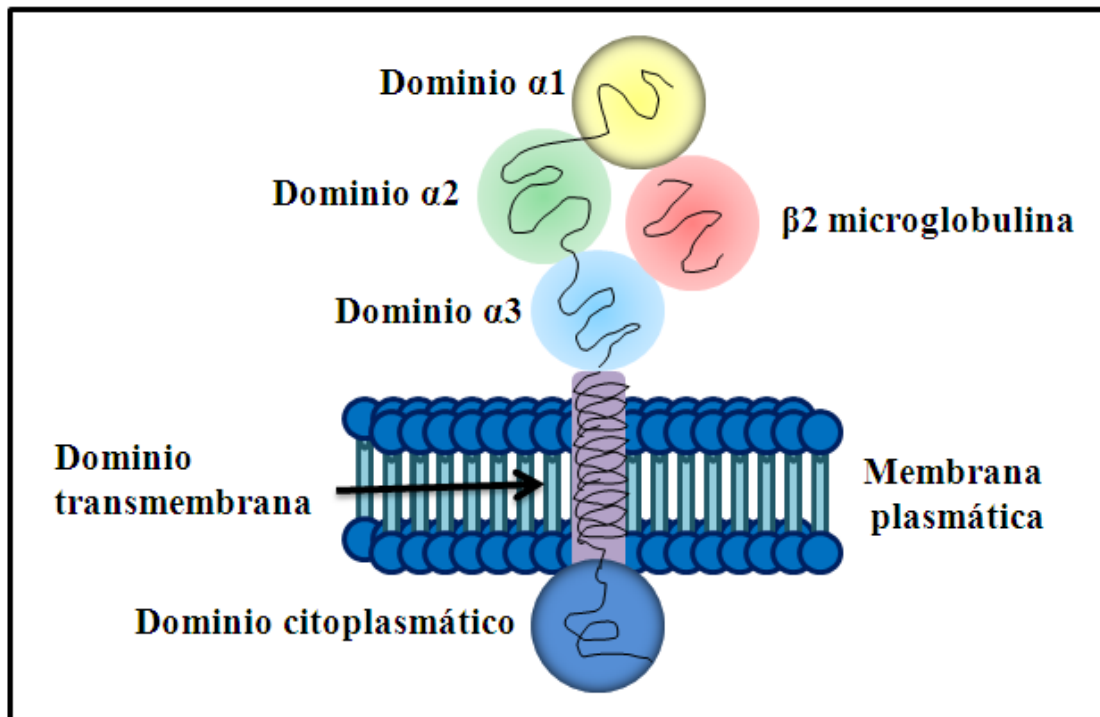


Figura I2: Esquema de una molécula de histocompatibilidad de clase I.

La estructura de las moléculas HLA de clase I fue descrita por primera vez por Bjorkman y sus colaboradores en 1987 mediante cristalografía. Concretamente se describió el antígeno HLA-A2 (Bjorkman et al. 1987a; Saper et al. 1991). Los cristales obtenidos revelaron la existencia de dos dominios tipo inmunoglobulina, próximos a la membrana ( $\alpha 3$  y  $\beta 2m$ ) y otros dos dominios más alejados ( $\alpha 1$  y  $\alpha 2$ ) asociados a un pequeño péptido no identificado ajeno a la molécula de histocompatibilidad.

Los dominios  $\alpha 1$  y  $\alpha 2$  presentaban una estructura compuesta por dos láminas  $\beta$  formadas cada una por cuatro hebras anti paralelas y una región de  $\alpha$ -hélice localizada en el extremo carboxi-terminal de las láminas  $\beta$ . Las láminas  $\beta$  de ambos dominios, forman la base de una valva flanqueadas por la  $\alpha$ -hélice orientadas hacia el exterior (Figura I3). Esta valva, constituye la estructura para la unión de los péptidos, (cuyo tamaño oscila entre los 8 y 11 aminoácidos (Rammensee et al. 1993)) y para la interacción con los receptores de las células T (TCR, del inglés *T Cell Receptors*) (Bjorkman et al. 1987b; Bjorkman et al. 1987a). La región más cercana a la membrana estaba formada por el dominio  $\alpha 3$  y la  $\beta 2M$ , cada uno de ellos formados por dos láminas  $\beta$  anti paralelas.



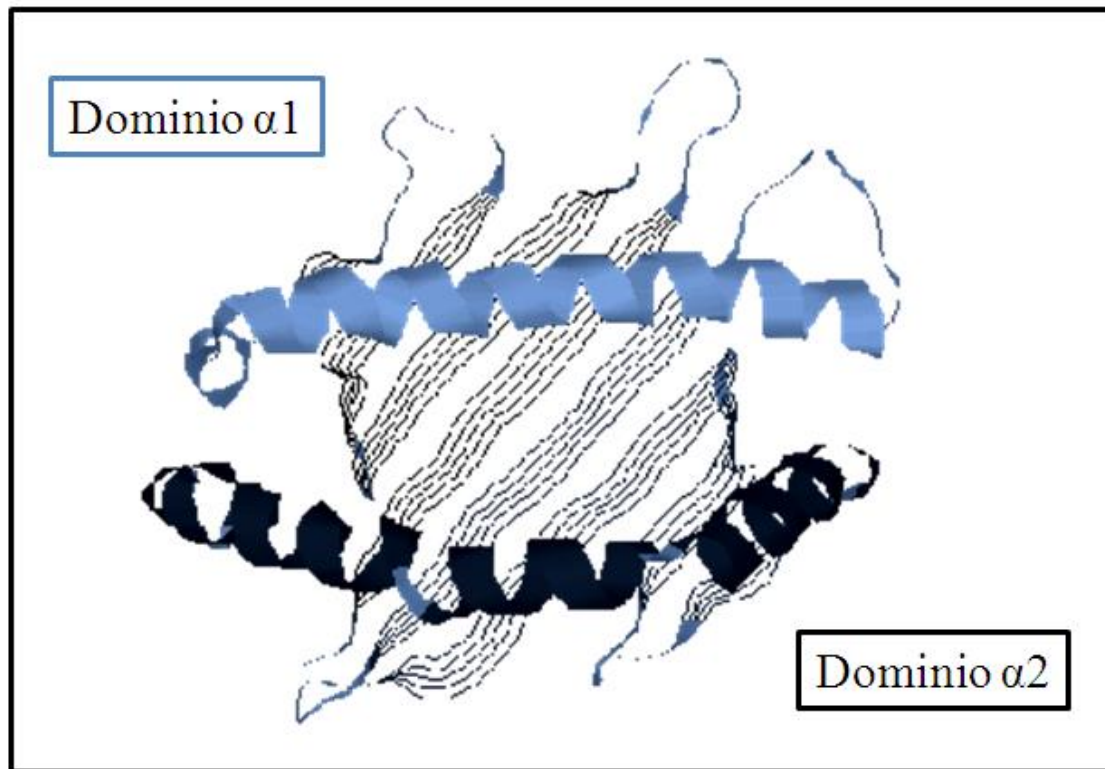


Figura I3: Estructura proteica de los dominios  $\alpha 1$  y  $\alpha 2$  de una molécula de Histocompatibilidad de clase I.

Las moléculas HLA de clase I no clásicas, por lo que se ha visto, tienen menos variabilidad en la valva de unión al péptido y por tanto su capacidad de unirse a éstos tiene una finalidad aún por determinar, si bien es cierto que son menos estudiadas que las moléculas clásicas, por lo que quizás esta falta de variabilidad se deba a la falta de estudios y análisis.

Las moléculas HLA de clase I clásicas, sin embargo, son muy polimórficas y su capacidad para unir péptidos está relacionada con la presentación antigénica, como se ha comentado anteriormente. El polimorfismo de estas moléculas se concentra en la valva formada por los dominios  $\alpha 1$  y  $\alpha 2$ , por lo que la variabilidad de los residuos que forman estos dominios es la que va a facilitar la unión de los distintos péptidos microbianos y con los receptores de las células T.

A finales de la década de los 80, Parham y colaboradores describieron 20 residuos aminoacídicos de alta variabilidad en las moléculas HLA (Parham et al. 1988). 16 de estos residuos están en contacto con el péptido antigénico (residuos 9, 24, 45, 62, 66, 67, 70, 74, 77, 80, 95, 97, 114, 116, 156 y 163 y tres de los 20 pueden estar en contacto con el TCR (residuos 65, 69 y 76) (Parham et al. 1988; Li & Bouvier 2004).



Saper y colaboradores, describieron una serie de depresiones en la valva de unión al péptido, denominados bolsillos de unión a péptidos ( en inglés, *Peptide-binding Pockets*) (Saper et al. 1991). En estos bolsillos se encuentran acomodadas una o más cadenas laterales del péptido. Los bolsillos hasta ahora identificados son 6, designados con las letras de la A la F. Los bolsillos B, C, D y E se localizan en las uniones de las láminas  $\beta$  con las dos  $\alpha$ -hélices, mientras que los bolsillos A y F se localizan en los extremos de las dos  $\alpha$ -hélices. Estando los residuos de estos dos últimos bolsillos más conservados que el resto (Barber & Parham 1993).

### **3.5 Función de las moléculas HLA de clase I**

Las moléculas HLA de clase I como se ha mencionado con anterioridad, poseen la capacidad de unir péptidos, los cuales pueden ser endógenos propios, tumorales y virales. De esta manera los hacen accesibles a la detección por parte de las células T que llevan en su superficie celular el marcador CD8 (mayoritariamente linfocitos citotóxicos). Esta función de unión y presentación antigénica, implica que cada molécula HLA de clase I puede unirse a un amplio espectro de péptidos. Sin embargo se ha comprobado que cada una de estas moléculas se une a un grupo determinado, lo cual indica que poseen selectividad alelo-específica (Falk et al. 1991). Los péptidos proceden principalmente de proteínas del interior celular degradadas en el proteosoma (Koopmann et al. 1997).

## **4. REGIÓN MHC DE CLASE II**

Esta región del MHC se localiza en la parte más centromérica del MHC en el cromosoma 6, comprendiendo aproximadamente 0.9 Mb (Horton et al. 2004). En ella se localizan los genes HLA de clase II clásicos (HLA-DP, -DQ y -DR) y no clásicos (HLA-DM y -DO) (ver Figura I1) (Trowsdale & Kelly 1985; Tonnellet et al. 1985).

### **4.1 Genes HLA de clase II clásicos**

Los genes HLA-DP, -DQ y -DR codifican para proteínas heterodiméricas, formadas por una cadena  $\alpha$  y una cadena  $\beta$ , que están implicadas en fenómenos de restricción del reconocimiento antigénico mediado por linfocitos T cooperadores (CD4+). Las cadenas  $\alpha$  de estas proteínas están codificadas por genes denominados “A”, tales como HLA-DPA, HLA-DQA y HLA-DRA, mientras que las cadenas  $\beta$  están

codificadas por genes denominados “B”, tales como, HLA-DPB, HLA-DQB y HLA-DRB.

Su distribución se centra fundamentalmente en células del sistema inmune encargadas de presentar antígenos, tales como, linfocitos B, macrófagos y células dendríticas, además de linfocitos T y células NK (Wu et al. 2013a).

### ***4.2 Genes HLA de clase II no clásicos***

Dentro de este grupo se encuentran los genes HLA-DM y HLA-DO (Trowsdale & Kelly 1985; Tonnelie et al. 1985). Estos genes al igual que los genes HLA de clase II clásicos, presentan genes A y B que codifican para las cadenas  $\alpha$  y  $\beta$  respectivamente. Los genes HLA-DM codifican para proteínas heterodiméricas implicadas en la unión de los péptidos del endosoma a las moléculas HLA de clase II clásicas (Sloan et al. 1995), mientras que la función de las proteínas codificadas por los genes HLA-DO posiblemente sea la de actuar como moduladores negativos de la presentación antigénica por parte de las moléculas HLA de clase II clásicas, mediante la interacción con la molécula de HLA-DM (Denzin et al. 1997).

Además de los genes HLA de clase II clásicos y no clásicos comentados anteriormente, existen en la región de clase II otros genes, como son: TAP 1 y TAP2, los cuales están implicados en la presentación antigénica de las moléculas HLA de clase I y codifican para proteínas transportadoras de péptidos citosólicos degradados (Trowsdale et al. 1990; Spies et al. 1990). Genes que codifican para subunidades del inmunoproteosoma que interviene en el procesamiento y degradación de los péptidos antigénicos (Glynne et al. 1991; Schmidt et al. 1999) y el gen HLA-Z, un pseudogen homólogo a los genes HLA de clase I (Beck et al. 1996).

### ***4.3 Estructura génica de los genes HLA de clase II***

Como se ha mencionado en apartados anteriores, las moléculas HLA de clase II están formadas por la unión de dos cadenas proteicas, una cadena  $\alpha$  y una cadena  $\beta$ , cada una codificada por un gen diferente.

Las cadenas  $\alpha$  se sintetizan a partir de los genes de tipo A, los cuales están formados por 5 exones (incluyendo la región 5'UT). **El exón 1** codifica para el péptido líder, **los exones 2 y 3** codifican para los dominios  $\alpha 1$  y  $\alpha 2$  respectivamente, **el exón 4** produce el péptido conector, la región transmembrana y el dominio citoplasmático. La región 3'UT esta codificada por parte del exón 4 y **el exón 5**.

Por otro lado, las cadenas  $\beta$  se sintetizan a partir de los genes de tipo B, los cuales están formados por 6 exones. **El exón 1** codifica para el péptido líder, **los exones 2 y 3** codifican para los dominios  $\beta 1$  y  $\beta 2$  respectivamente, el **exón 4** codifica para el péptido conector, el dominio transmembrana y parte del dominio citoplasmático. El resto del dominio citoplasmático estaría codificado por **el exón 5**, aunque esto puede variar en función de los diferentes genes de esta región.

#### 4.4 Moléculas HLA de clase II

Las moléculas HLA de clase II (HLA-DP, HLA-DQ, HLA-DR, HLA-DM y HLA-DO) son proteínas de membrana formadas por dos cadenas proteicas, una cadena  $\alpha$  de entre 33 y 35 kilodalton (Kd) y una cadena  $\beta$  de entre 26 y 28 kilodalton (Kd) ensambladas mediante uniones no covalentes y ancladas a la membrana plasmática. Cada una de las cadenas proteicas posee dos dominios extracelulares, (denominados  $\alpha 1$  y  $\alpha 2$  para la cadena pesada y  $\beta 1$  y  $\beta 2$  en el caso de la cadena ligera) un dominio transmembrana y una región citoplasmática. Los dominios  $\alpha 2$  y  $\beta 2$  son homólogos a las regiones constantes de las inmunoglobulinas (Figura I4).

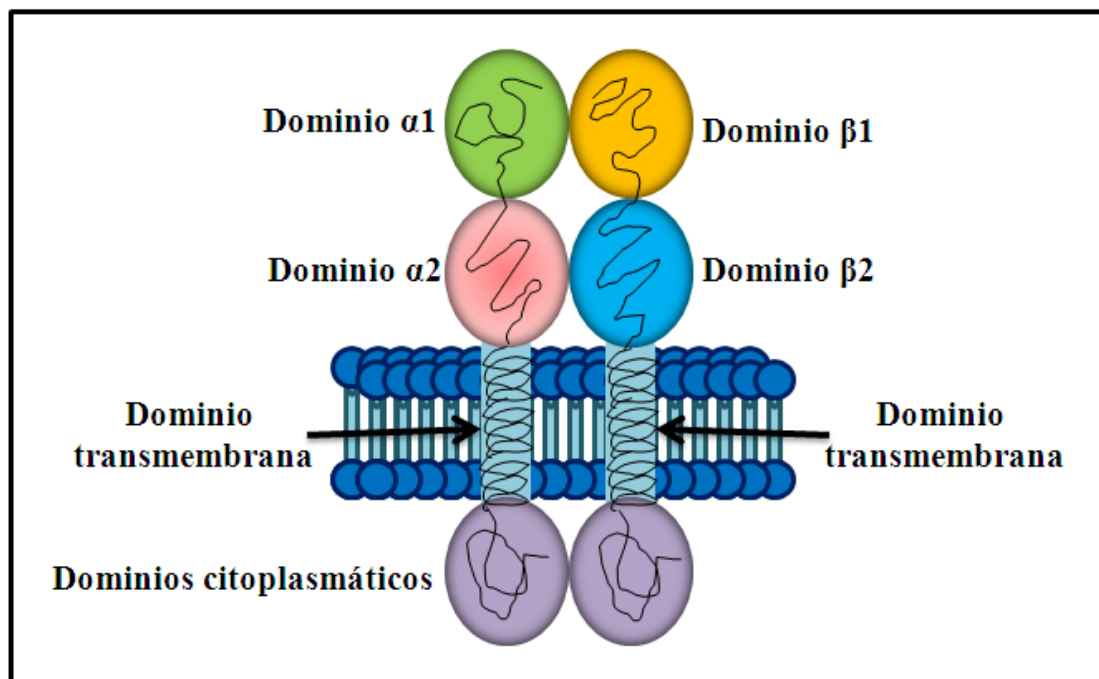


Figura I4: Esquema de una molécula de histocompatibilidad de clase II.

Las moléculas HLA de clase II se asocian en el interior de la célula con otra forma proteica denominada Cadena Invariante, la cual se codifica fuera del sistema HLA. Esta asociación se forma en el Retículo Endoplasmático y permite el transporte de las moléculas de clase II a los compartimentos endocíticos, donde se van a unir a los péptidos antigénicos. En estos compartimentos se va a degradar la Cadena Invariante en un conjunto de aminoácidos denominado CLIP (del inglés *Class II-associated Invariant Chain Peptides*) los cuales van a ocupar la valva de unión al antígeno (Ghosh et al. 1995). Finalmente, este complejo se disocia por la acción de las moléculas HLA-DM (Sloan et al. 1995) antes de su expresión en la superficie celular (Brocke et al. 2002).

La estructura tridimensional de las moléculas HLA de clase II fue determinada por Stern y colaboradores (Stern et al. 1994) gracias a las técnicas de difracción de rayos X. De esta forma, se observó que la estructura de estas moléculas es muy similar a las moléculas HLA de clase I descritas con anterioridad. Los dominios  $\alpha 1$  y  $\beta 1$  forman una valva muy similar a la de las moléculas HLA de clase I y también son capaces de unir péptidos, en este caso de mayor longitud. Esta valva consiste en una lámina  $\beta$  formada por ocho hebras anti paralelas, flanqueadas por dos regiones en  $\alpha$ -hélice. La valva de las moléculas HLA de clase II tiene la peculiaridad de estar abierta por sus extremos, lo que posibilita que pueda acomodar péptidos más grandes que la valva de las moléculas de clase I (Stern et al. 1994; Rammensee 1995).

### ***4.5 Función de las moléculas HLA de clase II***

Las moléculas HLA de clase II tienen la capacidad de unirse a un grupo heterogéneo de péptidos. A diferencia de las moléculas HLA de clase I, estos péptidos son en su mayor parte de origen extracelular, tales como, péptidos bacterianos y parasitarios. Las moléculas de clase II exponen estos péptidos en la superficie de las células presentadoras de antígenos para que puedan ser reconocidos por el TCR de los linfocitos T con marcador CD4 (mayoritariamente linfocitos T colaboradores).

La mayor parte de los péptidos presentados por las moléculas HLA de clase II provienen de la degradación endocítica de proteínas exógenas o de proteínas endógenas que acceden a los endosomas mediante la acción de proteasas como la cathepsina B (aunque este último caso se da en menor medida) (Katunuma et al. 1994).

La unión del péptido con la molécula HLA de clase II se da en unos compartimentos especiales denominados MIIC (del inglés, *MHC Class II Compartment*) (Tulp et al. 1994; Pieters 1997).

### 5. REGIÓN MHC DE CLASE III

Esta región del MHC está situada entre las regiones de clase I y de clase II, comprendiendo unas 0.7 Mb (Horton et al. 2004). Esta región es la más densa del genoma humano, por lo que los genes que la forman sufren un altísimo desequilibrio de ligamiento.

Los genes que se pueden encontrar en esta región, son genes que van a codificar para proteínas con función inmunológica, tales como genes TNF, genes del complemento de la vía clásica como C4A, C4B y C2, genes de la vía alternativa, genes del LY6, el gen del factor B del complemento y los genes HSP, los cuales codifican proteínas que intervienen en procesos de síntesis, plegamiento, ensamblaje, transporte y degradación de otras proteínas, aumentando su expresión en condiciones de estrés celular (ver Figura II).

### 6. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD EN EL POLLO

#### 6.1 Historia y definición

El sistema principal de histocompatibilidad del pollo (*Gallus gallus*) es uno de los más estudiados y mejor conocidos fuera del grupo de los mamíferos. Fue inicialmente descrito como parte de un sistema de grupos sanguíneos codificados en dos loci independientes (Briles & McGibbon 1948; Briles et al. 1950). Se denominó complejo B y fue el primer MHC identificado fuera de los mamíferos. Según avanzaban los estudios se observó la presencia de varios haplotipos, lo que evidenciaba que se trataba de un sistema bastante polimórfico (Briles & Briles 1982). Estudios moleculares corroboraron la presencia de este complejo B y de otra región de histocompatibilidad, independiente del complejo B a la que denominaron Rfp-Y (del inglés, *Restriction fragment pattern Y*) (Miller et al. 1994; Miller et al. 1996).

En 1999 Kaufman y colaboradores (Kaufman et al. 1999) obtuvieron el primer mapa completo del MHC del pollo a partir de pollos pertenecientes a la estirpe “Leghorn” (la más empleada en investigación). Sin embargo, debido a que esta estirpe descende de un número reducido de individuos (Simonson et al. 1989), se cuestiona si este mapa es representativo de la diversidad real de esta especie.

El material genético del pollo está distribuido en 39 parejas de cromosomas, 5 de gran tamaño, denominados macrocromosomas y 34 de pequeño tamaño, denominados

microcromosomas, de manera que podemos localizar la mitad de los genes en los macrocromosomas y la otra mitad en los microcromosomas (Guillemot et al. 1988). Es en el microcromosoma 16, uno de los más grandes (denominado GGA16) donde se localiza el MHC del pollo.

En la Figura I5 se pueden ver los principales grupos de genes que conforman el microcromosoma 16. El brazo corto del microcromosoma, denominado brazo p, es rico en Adeninas y Timinas. Este brazo aún no ha sido mapeado, por lo que no se conocen los genes que lo componen. Por otro lado, en el brazo largo, denominado brazo q, se han descrito varios grupos de genes. En la parte más cercana al centrómero existen múltiples copias de genes que codifican para receptores olfatorios y receptores “*scavenger*” (Miller et al. 2014). A continuación de estos genes y en dirección al telómero, está la única región de organización nucleolar de todo el genoma del pollo. Esta región ocupa entre 5 y 7 Mb del brazo q y contiene alrededor de 150 copias de los genes de ARN ribosómico 18S, 5.8S y 28S. Tras la región de organización nucleolar se encuentra la región MHC-Y, antes denominada región Rfp-Y. Distal a la región MHC-Y se encuentra una larga serie de repeticiones en tándem de 41 nucleótidos ricos en Guaninas y Citosinas (G+C) (Solinha et al. 2010). Tras esta serie de repeticiones, se encuentra la región MHC-B, la cual ha sido ampliamente estudiada por su papel en la presentación antigénica y en la resistencia a enfermedades. Por último, distal a la región MHC-B, se encuentran una serie de genes adicionales, entre los que destacan dos miembros de la familia CD1 que codifican para moléculas especializadas en la presentación antigénica.

Se han detectado en baja resolución otros genes dentro del microcromosoma 16, entre los cuales se pueden destacar el gen BLA, el cual codifica para la cadena  $\alpha$  de clase II (Kaufman et al. 1995; Salomonsen et al. 2003); el gen G9 (BAT8) que codifica una metiltransferasa en la región del complemento en otras especies (Spike et al. 1996) y el gen fB que codifica para el factor B del complemento (Kaufman et al. 1999).

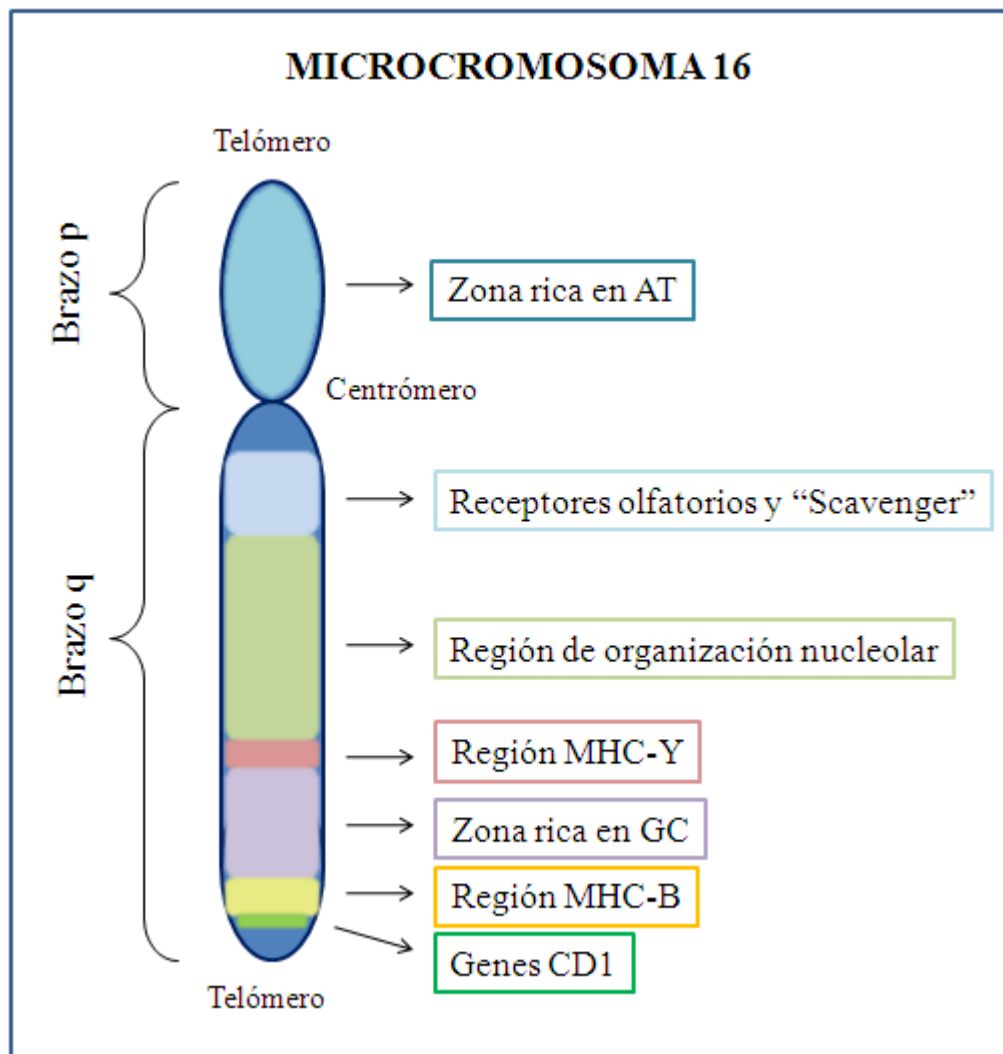


Figura I5: Mapa genético del microcromosoma 16 del pollo.

### 6.2 Región MHC-Y

Esta región, originariamente llamada región Rfp-Y, fue descubierta cuando se buscaban mediante sondas genéticas genes de clase I y II de la región MHC-B. Mediante RFLP se detectaron fragmentos de restricción que no pertenecían a la misma familia que los asignados a la región MHC-B (Briles et al. 1993). Tras este descubrimiento, dos genes MHC de clase I, dos genes MHC de clase II y un gen de tipo lectina tipo C, que previamente habían sido asignados a la región MHC-B (Guillemot & Auffray 1989), fueron reasignados a la región MHC-Y y localizados en el microcromosoma 16 (Miller et al. 1994; Miller et al. 1996). Todavía queda mucho por averiguar sobre esta región, pero los datos publicados actualmente, revelan la existencia de al menos 2 genes de tipo lectina tipo C, un gen MHC de clase I completo y un gen MHC de clase II $\beta$  truncado y una repetición de CR1 (Rogers et al. 2003). Los genes tipo

lectina detectados en esta región se han denominado *Ylec*. Son similares a los genes de mamíferos que codifican para receptores tipo lectina tipo C que guían la respuesta de las células *Natural Killer* en mamíferos (Iizuka et al. 2003).

En estudios anteriores se observó que los antígenos MHC-Y servían como aloantígenos débiles (Pharr et al. 1996), mientras que otros investigadores que trabajan con líneas celulares puras han demostrado que al menos uno de los genes MHC-Y de clase I es polimórfico y se expresa dinámicamente (Afanassieff et al. 2001; Hunt et al. 2006). La primera evidencia de la especialización de la valva de unión al antígeno se confirmó mediante estudios estructurales (Hee et al. 2009). Este estudio mostraba dentro de la valva de unión antigénica, ligandos no peptídicos (Hee et al. 2010). Lo que lleva a establecer que molécula MHC de clase I analizada en este estudio, (denominada YF1\*7.1), posee la estructura de una molécula MHC de clase I clásica que se une a ligandos no peptídicos y por lo tanto no clásicos. En humanos hay un ejemplo de este tipo de moléculas, la cual ha sido descrita en un estudio recientemente. Es el caso de la molécula MR1 humana, relacionada con el MHC de clase I. Esta molécula tiene una estructura similar a la clásica, sin embargo, presenta metabolitos de vitamina B de origen microbiano (Kjer-Nielsen et al. 2012). Estos resultados sugieren que las moléculas MHC-Y de clase I podrían contribuir a guiar respuestas inmunes de naturaleza especializada.

### **6.3 Región MHC-B**

Como se ha comentado anteriormente, los datos serológicos aportados por Briles en 1950, proporcionaron la primera evidencia de la presencia del MHC-B como un sistema aloantígeno desplegado en la superficie de los glóbulos rojos (Briles et al. 1950). Posteriormente, los investigadores Schierman y Nordskog (Schierman & Nordskog 1961) relacionaron el sistema del aloantígeno B con el complejo principal de histocompatibilidad del pollo al demostrar que los injertos de piel se rechazaban más rápidamente si el donante y el receptor no coincidían en su sistema B en comparación con otros aloantígenos sanguíneos. Más tarde Pink y colaboradores (Pink et al. 1977) definieron tres clases de antígenos MHC-B, cada uno de ellos con una distribución tisular diferente que indicaba la presencia de al menos tres loci MHC. Trabajos posteriores demostraron que la cooperación con las células del sistema inmune estaba restringida a los antígenos MHC-B (Maccubbin & Schierman 1986).



El primer mapa genético de la región MHC-B, se hizo en base al haplotipo B12 y abarcó múltiples genes de clase I y clase II y una parte de la región organizadora nucleolar (NOR) (Guillemot & Auffray 1989). Posteriormente y gracias al descubrimiento de la región MHC-Y (Briles et al. 1993) varios de los genes inicialmente asignados a la región MHC-B fueron reasignados a la región MHC-Y (Miller et al. 1994).

El mapa actual de la región MHC-B (Shiina et al. 2007; Salomonsen et al. 2014), comprende 46 genes distribuidos en 242 kb. Dentro de la región MHC-B se han definido 5 subregiones, las cuales se explican a continuación (Figura I6):

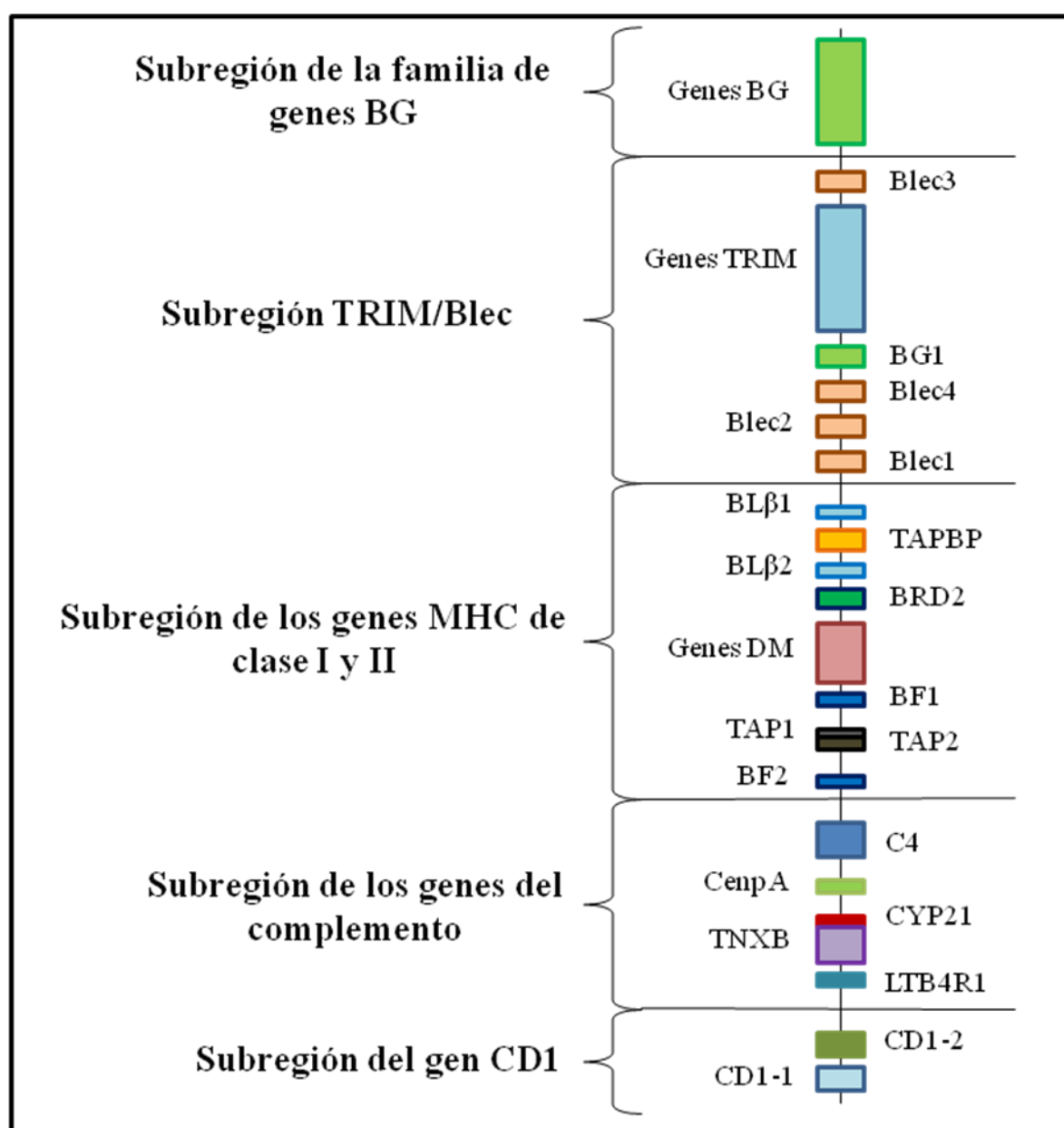


Figura I6: Mapa genético de la región MHC-B del pollo

### **Subregión de la familia de genes BG**

La mayoría de los genes de esta familia están agrupados en repeticiones en tándem en las cuales sus inserciones y deleciones (indels) de cuartetos de exones actúan como polimorfismos que permiten distinguir los diferentes haplotipos descritos (Salomonsen et al. 2014). Se trata de genes híbridos por lo que se sugiere que la recombinación contribuye a la diversidad de dichos haplotipos. Sólo dos genes están localizados fuera del grupo de repetición. El gen BG0, el cual se localiza en el cromosoma 2, fuera del MHC y el gen BG1, el cual se localiza a 115 kb del grupo de repetición, cerca de los genes MHC de clase I y II (ver Figura I6). Aparentemente todos los genes de la familia BG son polimórficos, aunque todavía se desconoce cómo contribuyen al sistema inmune. Sólo se ha podido relacionar el gen BG1 con la resistencia a la enfermedad de Marek (Goto et al. 2009).

Todas las moléculas BG poseen tres elementos estructurales destacables. Un dominio extracelular tipo inmunoglobulina del tipo región V que lleva una cisterna “extra” para el enlace disulfuro intermolecular; un dominio transmembrana tipo 1 y una cola citoplasmática compuesta de pequeños dominios (normalmente 7 aminoácidos), que forman estructuras de  $\alpha$ -hélice (Kaufman et al. 1990; Miller et al. 1991).

Algunas moléculas pertenecientes a la familia BG se expresan en eritrocitos, donde se pueden reconocer como aloantígenos que permiten distinguir serológicamente los haplotipos del sistema MHC-B. Otras moléculas se expresan en células epiteliales, particularmente en las del duodeno, el ciego y el hígado (Miller et al. 1990) y en una gran variedad de células importantes para la inmunidad, tales como células B, T y células dendríticas (Salomonsen et al. 2014).

### **Subregión TRIM/Blec**

En esta región se pueden encontrar un total de 25 genes, además del gen BG1 descrito en el apartado anterior. De estos 25 genes, se ha registrado la expresión de al menos 14 (Shiina et al. 2007). Se piensa que los genes TRIM están relacionados con la inmunidad innata proporcionando defensas frente a infecciones virales. Entre los genes del grupo *Blec*, destaca el gen *Blec2*, el cual se considera un posible receptor de células *natural killer*. Este gen además muestra un nivel alto de variabilidad y está bien expresado en tejidos linfoides (Miller & Taylor 2016).

### **Subregión de los genes MHC de clase I y clase II**

Esta región contiene genes involucrados en la presentación clásica de péptidos antigénicos. Por el momento se han descrito dos genes MHC de clase I, dos genes MHC de clase II $\beta$  y varios genes para el procesamiento y la carga de los antígenos (Miller & Taylor 2016).

Los dos genes MHC de clase I, denominados *BF1* y *BF2* codifican para moléculas MHC de clase I clásicas (Kaufman et al. 1999; Koch et al. 2007). Aunque estos dos loci parecen similares, funcionalmente son diferentes. *BF2* está implicado en la presentación del antígeno, mientras que *BF1* tiene aparentemente otra función principal distinta, ya que las moléculas codificadas por *BF1* pueden actuar de ligandos en interacciones con células NK (Miller & Taylor 2016). El gen *BF2* es el único descrito hasta el momento que está implicado en la presentación antigénica, lo que hace que el pollo sea un modelo experimental útil para investigar la especificidad del procesamiento, unión y presentación del antígeno y por tanto la respuesta inmune frente a diferentes enfermedades. Un buen ejemplo de este hecho es la evidencia de que *BF2* es una fuente de inmunidad protectora frente al virus del sarcoma de Rous (Hofmann et al. 2003).

De forma general, en la valva de unión al péptido de la molécula *BF2*, se dan las mismas interacciones moleculares que se producen en las valvas de otras moléculas MHC clásicas de clase I. Los péptidos antigénicos se encajan en la valva a través de interacciones entre los átomos del péptido y los residuos que forman dicha valva. Estos resultados se han confirmado mediante la determinación estructural de varias isoformas de la molécula *BF2* (Koch et al. 2007; Chappell et al. 2015). Al igual que con otras moléculas MHC clásicas de clase I que se unen al péptido, la unión de éstos se determina a través de interacciones preferenciales con los residuos que alinean los bolsillos de unión dispuestos a lo largo de la valva de la molécula *BF2*.

Además de los genes nombrados anteriormente, en esta subregión se pueden encontrar dos genes *TAP* (Transportadores asociados con el procesamiento de antígenos), los cuales codifican para moléculas transportadoras de ATP que envían péptidos del citoplasma celular al retículo endoplasmático (RE), donde son tomados por las moléculas MHC de clase I (Miller & Taylor 2016). También se puede encontrar el gen *TAPBP* que codifica para una tapasina, una proteína que media la interacción de las moléculas MHC de clase I con las moléculas *TAP* y facilita la carga del péptido.

Los genes MHC de clase II $\beta$  se localizan a ambos lados del gen de la tapasina (ver Figura I6). Uno de estos genes, se expresa más que el otro, por lo que se le considera como el gen MHC de clase II $\beta$  principal. En algunos haplotipos MHC-B se ha encontrado un tercer gen MHC de clase II $\beta$ , aunque todavía no se ha podido localizar (Jacob et al. 2000).

En esta región también se localizan los genes *DM*, los cuales codifican para moléculas que ayudan en la carga de las moléculas MHC de clase II. Estas moléculas de clase II permanecen dentro de las vesículas intracelulares y tienen un papel central en la carga de las moléculas MHC de clase II con el péptido antigénico. Se postula que los genes *DM*, *TAP* y *TAPBP* han co-evolucionado con los genes MHC de clase I y II.

Por último, en esta región se localiza el gen *BRD2* (también denominado *RING3*), el cual codifica para una proteína de unión a la cromatina y que aparentemente no tiene un papel directo en la inmunidad (Miller & Taylor 2016).

### **Subregión de los genes del complemento**

Esta región, comúnmente denominada como la región MHC de clase III, contiene sólo cinco genes (ver Figura I6). Cuatro de estos genes están muy conservados, el gen del factor del complemento *C4*; el gen *CenpA* que codifica para una proteína del centrómero; el gen *CYP21* que codifica para una hidroxilasa y el gen *TNXB* (tenascina XB). El quinto gen es el *LTB4R1*, el cual probablemente codifica para un receptor leucotrieno con un papel en inflamación. (Miller & Taylor 2016).

### **Subregión del gen *CD1***

En esta región se localizan dos genes *CD1*, *CD1-1* y *CD1-2*, a poca distancia de los dos genes MHC de clase I (ver Figura I6). Al contrario de lo que ocurre en mamíferos, estos genes están localizados en el pollo en el mismo cromosoma que el MHC. Las moléculas CD1 están muy estudiadas en mamíferos y son conocidas por contribuir a la respuesta inmune frente a bacterias. El estudio de la estructura de estas moléculas en el pollo ha puesto de manifiesto su relación con otros miembros de esta familia de genes y ha revelado la existencia de bolsillos de unión que no están presentes en otras moléculas CD1. El único bolsillo de la molécula CD2 es considerado como primordial ya que su valva de unión es muy pequeña y se cree que sólo puede acomodar en ella ácidos grasos o una cadena de lípidos de 16 carbonos (Zajonc et al. 2008). Por otro lado el gen *CD1* codifica para una molécula con una región de unión más grande de

doble bolsillo (Dvir et al. 2010), la cual es más similar a la molécula CD1 de los mamíferos en su capacidad para unir grandes lípidos.

### ***6.4 MHC y enfermedad en el pollo***

Tal y como ocurre en humanos, existen ciertos haplotipos del Sistema Principal de Histocompatibilidad del pollo que le confieren resistencia o susceptibilidad a ciertas enfermedades o dolencias. En cuanto a la influencia de los genes de la región MHC-Y frente a infecciones, los estudios realizados hasta el momento no desvelan una influencia clara o importante de estos genes frente a ciertas enfermedades. Se ha observado que en pollos con la enfermedad de Marek que eran homocigotos para un cierto haplotipo de MHC-Y, el riesgo de desarrollar tumores se incrementaba en 2, 3 veces más que en ausencia de dichos haplotipos. Sin embargo también se ha observado el caso contrario, animales que presentan menor virulencia de la enfermedad con menos tumores y que viven durante más tiempo (Pharr et al. 1996; Miller & Taylor 2016). Por lo que se necesitan más estudios para poder definir la relación entre los genes MHC-Y y la inmunidad en pollos. También se ha examinado la influencia de haplotipos MHC-Y en la evolución del Sarcoma de Rous. En pollos B2B5, se vio un efecto significativo sobre el patrón de crecimiento tumoral, el destino del tumor y las tasas de mortalidad (LePage et al. 2000). Es posible que los genes MHC-Y tengan un papel primordial en la defensa contra patógenos no virales, ya que se han encontrado genes MHC-Y de clase I al estudiar perfiles de genes que se activan con la inoculación de bacterias y glóbulos rojos de oveja en pollos (Connell et al. 2012; Geng et al. 2015; Wu et al. 2015).

La región MHC-B ha sido ampliamente estudiada para determinar su papel en la resistencia a enfermedades causadas por virus, bacterias y parásitos. Siendo las enfermedades más estudiadas, el Sarcoma de Rous y la enfermedad de Marek.

### **MHC-B y la enfermedad de Marek**

La enfermedad de Marek es una enfermedad muy contagiosa que afecta a los pollos. Se caracteriza por la presencia de linfomas de células T, así como la infiltración de linfocitos en nervios y órganos. La primera evidencia de la asociación entre MHC-B y la enfermedad de Marek, se describió en 1967 (Hansen et al. 1967), donde se evidenciaba el papel de dos haplotipos (B21 y B19) en la resistencia a esta enfermedad. Un trabajo posterior reveló como la virulencia de la enfermedad de Marek afectaba a la resistencia codificada por MHC-B. Otros estudios revelaron que los haplotipos MHC-B

también influían en la eficacia de las vacunas contra la enfermedad (Bacon & Witter 1994). En 1980 Elwood y Ruth Briles hicieron un gran avance en el estudio de la influencia de haplotipos MHC-B en la enfermedad de Marek al identificar haplotipos recombinantes (Briles & Briles 1980), sin embargo no lograron identificar el gen concreto responsable. Poco a poco se han ido realizando pequeños avances en el estudio de la influencia de los haplotipos recombinantes MHC-B en la resistencia a la enfermedad de Marek. Uno de estos avances se hizo gracias a la descripción de dos nuevos haplotipos recombinantes, BR2 y BR4, identificados como “idénticos” por los marcadores serológicos que había disponibles en el momento de su descubrimiento (Briles & Briles 1980). Posteriormente se analizaron estos dos haplotipos en dos líneas congénitas en desarrollo para determinar si tenían fenotipos idénticos respecto a la resistencia a la enfermedad (Schat et al. 1994). Los resultados obtenidos mostraron como las dos líneas diferían en cuanto a su resistencia a la enfermedad, siendo una de ellas más vulnerable a la misma. Posteriormente se mapearon ambos haplotipos, lo que indicó que poseían alelos BG1, denominados BG1\*R4 y BG1\*R2, los cuales eran completamente idénticos en su secuencia codificante, pero diferentes en la región 3'UTR (del inglés, *3' untranslated region*). Por el momento no se entienden bien las consecuencias de esta diferencia, se postula que podrían alterar la actividad de BG1 y con ello la incidencia sobre la enfermedad. Con estos hallazgos se muestra que gran parte de la influencia genética que posee MHC-B sobre la enfermedad de Marek podría ser atribuida a BG1 (Miller & Taylor 2016).

### **MHC-B y el virus del Sarcoma de Rous**

El virus del Sarcoma de Rous es un retrovirus que causa tumores en pollos. La aparición de estos tumores está mediada por el oncogen *v-src* (Brugge & Erikson 1977). La influencia de MHC-B sobre el virus del Sarcoma de Rous se describió por primera vez en dos publicaciones de manera simultánea (Collins et al. 1977; Schierman et al. 1977). En estos estudios cruzaron animales con diferentes haplotipos de MHC-B y los expusieron al virus. Lo que observaron fue que ciertos haplotipos provocaban que los tumores evolucionaran de manera diferente, disminuyendo el crecimiento en algunos casos hasta en un 94% (pollos heterocigotos para *B6B13*) (Schierman et al. 1977).

Varios estudios se han centrado en encontrar el gen o genes dentro de la región MHC-B, que determina la regresión de los tumores provocados por el virus del Sarcoma de Rous. Estudios antiguos centrados en el haplotipo recombinante BG4-BF12,

observaron que la resistencia al virus se localizaba en la porción BF del MHC-B (Plachy & Benda 1981). Otros estudios más recientes han visto que el crecimiento de tumores difiere entre líneas de pollos con idénticos genes MHC y diferencias en el genoma, lo que indicaría un papel importante para genes no MHC. Sin embargo, las metástasis por tumores provocados por el virus difieren en cuanto a diferencias en genes no MHC así como en genes MHC-B (Taylor 2004).

### **7. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD EN OTRAS AVES**

Como se ha comentado en apartados anteriores, existen dos clases de genes MHC, los genes MHC de clase I y los genes MHC de clase II, cuyas funciones son de sobra conocidas. Los genes MHC de clase I se ocupan principalmente de infecciones intracelulares, como por ejemplo las provocadas por virus, mientras que los genes MHC de clase II se ocupan de las infecciones extracelulares, como las provocadas por bacterias (Abbas et al. 1994; Hughes & Yeager 1998). Hay algunas infecciones que provocan la activación de las dos clases de genes, como es el caso de la malaria. En el apartado 6, se ha comentado el escaso número de genes MHC que posee el pollo. Existen otras aves en las que se han descrito pocos genes MHC, como es el caso del faisán de cuello anillado (*Phasianus colchicus*), el cual posee dos genes MHC de clase I y dos genes MHC de clase II y el urogallo negro (*Tetrao tetrix*) que también tiene un número muy bajo de genes MHC de clase I y II (Wittzell et al. 1999; Strand et al. 2007). Estas tres aves *Galliformes*, presentan un número bajo de genes MHC de clase II, al igual que el pingüino (Bollmer et al. 2007). Sin embargo a excepción de estos dos órdenes de aves, la mayoría de las aves estudiadas hasta la fecha presentan un mayor número de genes MHC de clase I y II (Hess & Edwards 2002; Ekblom et al. 2003). Este factor es particularmente destacable en el caso de las aves *Passeriformes* (Bonneaud et al. 2004; Miller & Lambert 2004; Richardson & Westerdahl 2003).

#### **7.1 Genes MHC en aves *Passeriformes***

Durante la última década se han secuenciado parcial o totalmente los genes MHC de clase I de varias especies de aves *Passeriformes*, tales como: *Acrocephalus arundinaceus* (Westerdahl 1999), *Carpodacus erythrinus* (Promerova 2009), *Passer*



*domesticus* (Bonneaud 2004), *Taeniopygia guttata* (Balakrishnan 2010, Ekblom 2010) y *Cyanistes caeruleus* (Schout 2011). Estas aves tienen entre 2 y 8 genes MHC de clase I, aunque varios estudios basados en RFLPs sugieren que los *Passeriformes* podrían tener loci adicionales (Wittzell 1998, Westerdahl 2007). Por lo tanto, las aves *Passeriformes* poseen un número mayor de genes MHC comparado con el pollo, además de que sus genes MHC y los intrones que estos contienen son más grandes (Westerdahl et al. 2004). Actualmente no está claro si existe un locus MHC-Y en aves *Passeriformes*, aunque si se han descrito alelos MHC con una baja variabilidad (Jarvi et al. 2004). El gorrión común (*House sparrow*), por ejemplo, posee dos grupos de alelos transcritos para MHC de clase I, uno de ellos muy variable y otro menos variable (Bonneaud et al. 2004). Determinar a qué locus pertenecen los genes MHC de las aves *Passeriformes* es bastante complejo y difícil. La alta similitud entre alelos pertenecientes a loci diferentes es probablemente resultado de la conversión génica. Este proceso se ha observado tanto dentro de un mismo locus como entre varios loci en varias especies, como por ejemplo en el faisán de cuello anillado (Wittzell et al. 1999).

Sin embargo, aunque es bastante difícil, se ha conseguido llevar a cabo la amplificación específica de locus en ciertas especies, como es el caso del mirlo de alas rojas (Gasper et al. 2001). Todo parece indicar que el mejor camino para obtener el mapa completo del MHC de estas aves es la secuenciación de su genoma, como es el caso del proyecto del genoma del diamante mandarín (*Taeniopygia guttata*).

Cuando un gran número de genes MHC se mantienen dentro de una población durante largos periodos de tiempo, se debe probablemente a la selección de balance (Hedrick 2002). La evolución trans específica sin embargo, sugiere que los alelos MHC no sólo se mantienen dentro de una misma población, sino que sobreviven a varios eventos de especiación. Para ello estos alelos tienen que ser idénticos al menos en las regiones que codifican para la zona de unión al péptido de la molécula de MHC y que además el árbol basado en los genes MHC de varias especies diferentes, esté en discordancia con la filogenia real de esas especies (Edwards & Hedrick 1998), (basado por ejemplo en el gen del citocromo b mitocondrial). Las fuerzas más frecuentes detrás de la selección de balance son la selección negativa dependiente de frecuencia, la cual se da cuando los alelos denominados “raros” tienen ventaja sobre los más comunes (Bodmer 1972). Y la ventaja por heterocigosis, la cual se da cuando un individuo con dos alelos diferentes está en ventaja sobre un individuo que posee un solo alelo (Doherty & Zinkernagel 1975).



Una presión de selección de un único patógeno que se comparte entre dos especies divergentes, como pueden ser el carricero tordal (*Acrocephalus arundinaceus*) y el gorrión común (*Passer domesticus*) podría contrarrestar el proceso de divergencia e incluso mantener alelos MHC idénticos en estas dos especies distantemente relacionadas. Estas dos aves, por ejemplo, son infectadas por la misma cepa de malaria aviar (*Plasmodium relictum* GRW4), por lo que podrían estar sometidas a presiones de selección similares para este tipo de infección. Sin embargo todavía no se han reportado alelos MHC idénticos entre el carricero tordal y el gorrión común (Westerdahl 2007).

### **7.2 MHC y enfermedad en aves *Passeriformes***

La alta diversidad de genes MHC se mantiene probablemente por la selección de un gran número de patógenos. Como se ha comentado en el apartado anterior, la selección negativa dependiente de frecuencia podría mantener esta diversidad, pero entonces la prevalencia de patógenos debería variar en el tiempo y/o en el espacio (Hedrick 2002). Diferentes especies de malaria aviar (*Plasmodium*, *Haemoproteus* y *Leucocytozoon*) podrían ser uno de los conjuntos de patógenos que mantienen la diversidad genética en las aves *Passeriformes*, debido a que son patógenos comunes, pueden ser letales y su prevalencia varía en el tiempo y en el espacio (Atkinson & van Riper 1991; Bensch et al. 2000; Ricklefs & Fallon 2002; Bensch & Akesson 2003).

Se han realizado varios estudios basados en el gorrión común (*Passer domesticus*) y en el carricero tordal (*Acrocephalus arundinaceus*), en los cuales se ha analizado la posible asociación entre la prevalencia de infecciones por malaria y la presencia de ciertos alelos MHC de clase I. La prevalencia de la infección por malaria se analizó mediante la amplificación del gen del citocromo b usando la técnica de la reacción en cadena de la polimerasa (PCR) de los parásitos de la malaria (géneros *Plasmodium* y *Haemoproteus*), a partir de sangre que se extrajo a las aves de estudio. Bonneaud y colaboradores (Bonneaud et al. 2006), investigaron estas asociaciones entre malaria y alelos MHC de clase I en dos poblaciones de gorrión común (*Passer domesticus*), encontrando que la prevalencia de la infección por malaria (SGS1) era bastante elevada, de un 68% y un 56 %. Además encontraron una asociación negativa entre el alelo MHC a151 y la infección por malaria (SGS1) en una de las poblaciones y otra entre el alelo MHC a172 y SGS1 en la otra población de estudio.

Por otro lado Westerdahl y colaboradores (Westerdahl et al. 2005), investigaron la asociación entre 3 infecciones diferentes de malaria (GRW1, GRW2 y GRW3) y 23

alelos MHC de clase I en el carricero tordal (*Acrocephalus arundinaceus*). Encontraron una asociación positiva entre el alelo MHC B4b y la infección GRW2. La prevalencia de esta infección en la población de estudio era baja, sólo un 8% de los individuos la tenían, además según los análisis realizados se sugiere que este tipo de infección es severa, incluso letal en esta especie de ave.

Las diferencias encontradas en los dos estudios anteriores, observándose una asociación negativa en el gorrión común y una asociación positiva en el carricero tordal, pueden deberse a la diferente virulencia de las cepas de malaria empleadas en cada uno de los estudios, ya que en el caso del gorrión común, la infección era leve o benigna, mientras que en el caso del carricero tordal la infección era más severa incluso letal. De tal forma que, un alelo MHC puede proporcionar resistencia a una infección por malaria, dando total protección, o disminuyendo el efecto de una infección letal, haciéndola más leve (Westerdahl 2007).

## 8. INTRONES

### 8.1 Historia y definición

Un intrón es un segmento de ADN localizado dentro de un gen que se transcribe, pero que posteriormente es eliminado en la maduración del ARN.

Los intrones fueron descubiertos en 1977 y se localizan principalmente en organismos eucariotas, donde forman una parte importante del genoma (Gilson et al. 2006). La densidad de los intrones oscila entre unos pocos en el genoma completo de algunos protistas (Mair et al. 2000) a una media de ocho intrones por gen en el caso de los humanos (Sakharkar et al. 2004). La eliminación de los intrones se produce mediante un mecanismo de corte y empalme (en inglés *splicing*). Los sitios canónicos de *splicing* son secuencias cortas muy conservadas que se localizan en los límites de unión de los intrones con los exones. Estas secuencias son reconocidas por pequeñas moléculas de ARN nuclear que interactúan con una serie de proteínas formando partículas ribonucleoproteicas que constituyen el espliceosoma. Este sistema corta la secuencia génica por los sitios canónicos, uniendo los exones y eliminando los intrones.

En la actualidad existen dos teorías sobre el origen de estos segmentos génicos tan peculiares, los intrones. La teoría de los intrones tempranos y la teoría de los intrones tardíos. Según la primera teoría, el ancestro común de los organismos eucariotas y procariotas poseía en su genoma una gran mezcla de intrones, que facilitó la creación de genes tempranos mediante el barajeo de los exones y que la ausencia de

intrones en los organismos procariotas es una consecuencia secundaria de la evolución natural de los mismos en los que se priorizan cortos periodos reproductivos y de vida (Lynch 2002). Al perder los intrones los procariotas también perdieron la capacidad de evolucionar de manera más compleja, debido a que los exones no sufrían tanta recombinación que pudiera originar nuevos genes. La teoría de los intrones tardíos por otro lado, postula que la gran mayoría de los intrones surgieron dentro de organismos eucariotas multicelulares, insertándose más o menos al azar en los genes preexistentes, aunque no se descarta que tuvieran un papel posterior en la evolución adaptativa de las proteínas (Lynch 2002). Las visiones extremas de los seguidores de ambas teorías se han ablandado en parte, aunque todavía existe una gran controversia sobre la biología evolutiva de los intrones. La idea de que la pérdida paralela masiva de intrones ocurrió en todos los linajes ancestrales de los procariotas de hoy en día plantea obvias dificultades lógicas y no aborda por qué la reducción el tamaño del genoma no era una prioridad anterior en la evolución. Por otro lado, el origen temprano de los intrones ya no puede ser negado, ya que se han detectado intrones en procariotas.

### ***8.2 Tipos de intrones***

Como se ha comentado en el apartado anterior, fue en 1977 cuando se descubrieron los intrones. Fue tras aplicar la secuenciación a genes eucariontes y ver como muchos de estos genes contenían una serie de secuencias interpuestas que separaban diferentes segmentos de ADN codificante. Actualmente se reconocen ocho tipos distintos de intrones en organismos eucariotas y otras formas en arqueobacterias (Brown 2006).

#### **Intrones GU-AG**

Son aquellos intrones cuyas secuencias consenso de corte y empalme comienzan con GU en el extremo 5' y acaban con AG en el extremo 3'. Se localizan en el pre ARN mensajero eucarionte. Todos los intrones pertenecientes a este grupo se cortan y empalman de la misma forma. En primer lugar se produce la degradación del sitio de corte y empalme 5' por una reacción de transesterificación promovida por los grupos oxhidrilo unidos al carbono 2' de un nucleótido de adenosina localizado dentro de la secuencia del propio intrón. El resultado de esta reacción es la degradación del enlace fosfodiéster en el sitio de corte y empalme y la formación de un nuevo enlace fosfodiéster 5'-2' que une el primer nucleótido del intrón con la adenosina interior. Lo

que provoca que el intrón forme un bucle sobre sí mismo formando una estructura de lazo (Figura I7). Por último se produce la degradación del sitio de corte y empalme 3' tras otra reacción de transesterificación, promovida por el grupo 3' OH unido al extremo del exón. Este grupo ataca el enlace fosfodiéster en el extremo 3' del intrón, lo degrada y libera para su posterior degradación. Al mismo tiempo los exones que estaban a ambos lados del intrón se unen completando así el proceso de corte y empalme (Figura I7) (Brown 2006).

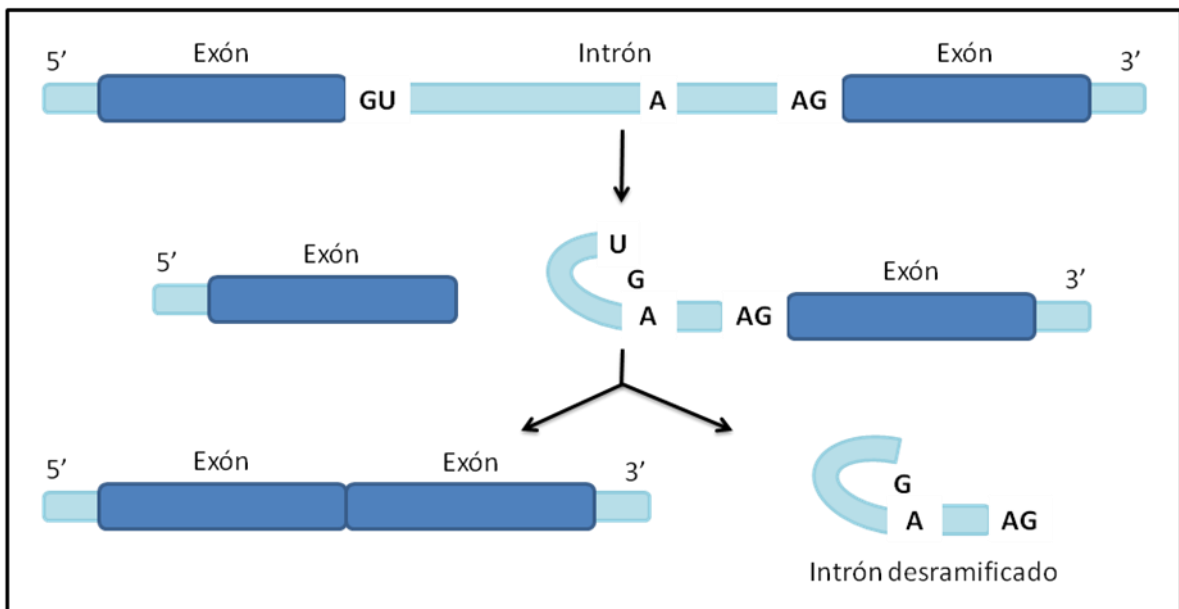


Figura I7: Esquema del proceso de corte y empalme de los intrones GU-AG.

### Intrones AU-AC

Estos intrones se han descrito en alrededor de 20 genes de organismos tan diversos como seres humanos, plantas y *Drosophila*. Al igual que los intrones GU-AG se localizan en el pre ARN mensajero eucarionte. Sus secuencias consenso de corte y empalme comienzan con AU en el extremo 5' y terminan con AC en el extremo 3'. Además de poseer estas secuencias consenso, tienen una secuencia conservada en el sitio de ramificación con el consenso 5'-UCCUUAAC-3', cuya última adenosina participa en la primera reacción de transesterificación del mecanismo de corte y empalme. Este mecanismo es similar al que se produce en los intrones GU-AG, la diferencia radica en los grupos del espliceosoma que participan en el proceso (Brown 2006).

### **Intrones del Grupo I**

En este grupo se engloban aquellos intrones que se localizan en Pre-rARN nuclear eucarionte, como es el caso del eucarionte microbiano *Tetrahymena*, en ARN de orgánulos como mitocondrias y cloroplastos y en algunos ARN microbianos, como la cianobacteria *Anabaena*. La vía de corte y empalme de este tipo de intrones es similar a la de los intrones de tipo GU-AG y AU-AC, ya que se producen dos reacciones de transesterificación. La primera reacción la provoca un nucleótido libre o un nucleósido y no un nucleótido que se sitúe dentro del intrón como en casos anteriores. Tras la degradación del sitio de corte y empalme 5', una segunda reacción de transesterificación provoca la ruptura del enlace fosfodiéster del sitio 3', causando su degradación, la unión de los dos exones y la liberación del intrón. La característica principal de la vía de corte y empalme de los intrones pertenecientes al grupo I es que tiene lugar en ausencia de proteínas y por tanto es autocatalítica, ya que es el propio ARN el que tiene actividad enzimática. Este fue el primer ejemplo de enzima de ARN o ribozima en ser descubierto. Posteriormente se vieron más ejemplos (Brown 2006).

### **Intrones de pre-tARN**

Los intrones del ARN de transferencia son relativamente comunes en los eucariontes inferiores y menos frecuentes en vertebrados. En humanos, sólo el 6% de los genes de tARN contienen intrones. A diferencia de otros tipos de intrones, el corte y empalme de los intrones de tARN no implica reacciones de transesterificación. Los dos sitios de corte y empalme son cortados por una endonucleasa. Esta enzima contiene cuatro subunidades no idénticas. Una de ellas identifica los sitios de corte y empalme, otras dos practican los cortes en dichos sitios, lo que deja una estructura de fosfato cíclico unida al extremo 3' del exón corriente arriba y un grupo oxhidrilo en el extremo 5' del exón corriente abajo. Posteriormente el fosfato cíclico es convertido en un 3'-OH por una fosfodiesterasa y el extremo 5'OH es convertido en 5'-P por una cinasa. Estos dos extremos se ligan finalmente por una ligasa. Es decir, una sola proteína aporta las funciones de fosfodiesterasa, cinasa y ligasa (Brown 2006).

### **Intrones del grupo II**

Estos intrones se localizan en orgánulos de plantas y hongos, tanto en pre-mARN como en pre-rARN. También se han descrito unos pocos en procariotas. Tienen una estructura secundaria muy característica y son capaces de autocortarse y empalmarse en el tubo de ensayo, aunque son distintos a los intrones del grupo I. El

mecanismo de corte y empalme está estrechamente relacionado con el de los intrones GU-AG, ya que la primera transesterificación es promovida por el grupo oxhidrilo de un nucleótido de adenosina interno y el intrón acaba formando una estructura de lazo. Esta similitud sugiere que ambos tipos de intrones pueden tener un origen evolutivo común (Brown 2006).

### **Intrones del grupo III**

Estos intrones también se encuentran en los genomas de algunos orgánulos y son capaces de autocortarse y empalmarse mediante un mecanismo muy similar al de los intrones del grupo II. Sin embargo son más pequeños y su estructura secundaria es muy peculiar (Brown 2006).

### **Intrones gemelos**

Se trata de estructuras formadas por dos o más intrones del grupo II o del grupo III. Los intrones gemelos más simples constan de un intrón sepultado en otro, mientras que los más complejos contienen múltiples intrones sepultados. En general cada intrón que forma un intrón gemelo es cortado y empalmado en una secuencia definida (Brown 2006).

### **Intrones arqueobacterianos**

Estos intrones se localizan en los genes de ARNt y ARNr. Son eliminados mediante una ribonucleasa similar a la que participa en el proceso de corte y empalme del pre-tARN eucarionte.

## ***8.3 La función de los intrones***

La formación exón-intrón de muchos genes eucariotas plantea la pregunta de si esta organización única tiene alguna función o es simplemente el resultado de la propagación de secuencias intrónicas sin función en los genomas eucariotas (Chorev & Carmel 2012).

Se cree que la presencia de intrones en un genoma impone una carga sustancial sobre el huésped. En primer lugar a diferencia de los intrones que se auto cortan (denominados intrones *self-splicing*), los que son eliminados por el espliceosoma, necesitan este sistema, el cual está entre los mayores complejos moleculares de la célula, comprendiendo 5 snRNAs y más de 150 proteínas (Wahl et al. 2009). Por lo que

los genomas portadores de intrones, necesitan codificar todas las proteínas y snRNAs necesarios para llevar a cabo el proceso de corte y empalme (*splicing*). En segundo lugar, la transcripción de los intrones es costosa en términos de tiempo y energía, ya que la velocidad media de elongación de una ARN polimerasa es de 60 bases por segundo (Singh & Padgett 2009) lo que significa que la transcripción de algunos intrones largos puede llevar horas. En tercer lugar, el reconocimiento de las secuencias de unión por parte del espliceosoma está dirigido por una serie de elementos reguladores en *cis*. Esto hace al organismo vulnerable a mutaciones sinónimas que de otro modo no tendrían un efecto notable, de hecho se estima que más del 50% de los trastornos genéticos en humanos están causados por la interrupción del patrón de corte y empalme (Lopez-Bigas et al. 2005) Finalmente, el mal funcionamiento de cualquier proteína o snRNA que forma parte del espliceosoma tendrá un efecto perjudicial sobre la célula (Chorev & Carmel 2012).

Un buen ejemplo que muestra que los intrones ejercen una función importante en organismos eucariotas es el incremento de la abundancia de proteínas en genes portadores de intrones. Este efecto fue observado por primera vez en construcciones del virus 40 vacuolado del simio, cuyo producto proteico se hizo indetectable una vez se eliminaron todos sus intrones (Gruss et al. 1979). Usando construcciones virales similares, se demostró que la eliminación de los intrones también afecta a un nivel de ARNm. En algunos casos las construcciones portadoras de intrones expresaban hasta 400 veces más que las que carecían de ellos (Buchman & Berg 1988).

Estudios posteriores describieron el mismo fenómeno asociado a otros intrones en muchas especies de eucariotas, sugiriendo que esta función abarca un amplio rango (Le Hir et al. 2003). Un ejemplo de ello se da en las plantas, donde este efecto intrónico se denomina “mejora mediada por intrón” y está ampliamente descrito (Mascarenhas et al. 1990; Luehrsen & Walbot 1991; Akua et al. 2010). En levaduras, se observó que genes portadores de intrones producían más ARNm y más proteínas que los genes carentes de los mismos (Juneau et al. 2006). En mamíferos también se ha observado este fenómeno, donde los genes portadores de intrones tenían una expresión más abundante y amplia que los no portadores (Shabalina et al. 2010).

No existe un único mecanismo mediante el cual los intrones mejoran la expresión de los genes que los portan. En algunos casos el mecanismo de actuación no se conoce todavía, sin embargo, en aquellos casos donde sí se ha podido describir,

parece que los intrones son capaces de afectar cada uno de los pasos de la maduración del ARN mensajero (Chorev & Carmel 2012).

Se han descrito una gran variedad de funciones asociadas a los intrones (Chorev & Carmel 2012). Estas funciones se han observado en las diferentes fases por las que pasa un intrón, partiendo de la secuencia de ADN del mismo, hasta que este es liberado por el espliceosoma. A continuación se describen algunos ejemplos de funciones que poseen los intrones:

### **Funciones asociadas al intrón genómico**

Los intrones pueden ser tomados como una serie de secuencias de libre selección que se encuentran dentro de los genes. Desde un punto de vista evolutivo, esta característica hace que casi cualquier mutación en la secuencia intrónica sea tolerable. Los intrones por tanto, tienen el potencial de servir como repositorios de elementos *cis*, participando en la regulación de la transcripción y la organización del genoma (Chorev & Carmel 2012).

Varios estudios han identificado elementos de ADN reguladores del inicio de la transcripción dentro de secuencias intrónicas. Estos elementos incluyen potenciadores (Tourmente et al. 1993), silenciadores (Tourmente et al. 1993) y otros elementos que modulan la función del promotor principal (Bornstein et al. 1988). En la mayoría de los casos, estos elementos reguladores se encuentran en los primeros intrones de cada gen (Bornstein et al. 1988).

Algunos intrones no albergan elementos que modifican la eficiencia del promotor principal, sino que alojan un promotor alternativo que da lugar, cuando se activa, a una isoforma con un sitio de transcripción diferente (Schoy et al. 2000). Otros casos que se han estudiado, muestran como algunos intrones pueden influir en la finalización del proceso de transcripción. Un buen ejemplo de esto es el caso del intrón 2 del gen  $\beta$ -globina humano, el cual reduce la eficiencia de la formación del extremo 3' si es eliminado o sustituido por otro intrón (Antoniou et al. 1998).

### **Funciones asociadas al intrón transcrito**

Los intrones pasan por el proceso de transcripción al igual que los exones, para formar el pre-ARN mensajero. Usando varias técnicas se ha estimado el ratio de elongación de la ARN polimerasa II (Ardehali & Lis 2009). La medida reciente de este ratio en diferentes regiones de nueve genes humanos grandes, muestra un ratio



homogéneo de 3.8 kb por minuto (Singh & Padgett 2009). Por lo que muchos intrones requieren minutos, horas o incluso días para transcribirse. Este hecho, plantea la posibilidad de que los intrones puedan servir como herramientas de manejo de los tiempos entre la activación de un gen y la aparición de su proteína (Gubb 1986).

Este factor se ha demostrado en el gen E74 que da comienzo a la metamorfosis de *Drosophila melanogaster*. El gen está compuesto por tres transcritos, siendo el primero de ellos de 60 kb de longitud, antes del proceso de *splicing* y de 6 kb después. Este gen está inducido por una hormona esteroide y aparece en el citoplasma aproximadamente una hora después de su inducción. La medida del ratio de la polimerasa en este gen fue de 1.1 kb por minuto (Thummel et al. 1990), lo que sugiere que es el tiempo de transcripción de los intrones lo que provoca el aumento de tiempo.

### **Funciones asociadas al “intrón cortado”**

El proceso de corte y empalme del pre-ARN mensajero es llevado a cabo por el espliceosoma, el cual está formado por cinco pequeños ARN nucleares principales (snARNs), U1, U2, U4, U5 y U6, algunas proteínas nucleares y otras proteínas satélite (Wahl et al. 2009).

Varios estudios muestran que el proceso de corte y empalme de la mayoría de los intrones ocurre concomitantemente con el de transcripción y que estos dos procesos celulares están fuertemente acoplados, principalmente a través del dominio carboxi terminal de la ARN polimerasa II (Khodor et al. 2011).

Algunos sitios de corte y empalme (en inglés *splice sites*) son reconocidos como tales por el espliceosoma, mientras que otros sitios, dependiendo de los tejidos, el tiempo y las circunstancias, pueden ser omitidos por el espliceosoma y dar lugar a un proceso de corte y empalme alternativo. Este proceso, permite la diversidad del proteoma, que excede mucho el número de genes en el genoma (Nilsen & Graveley 2010). Un ejemplo destacado de este fenómeno es el gen *Dscam* de *Drosophila melanogaster*, el cual tiene el potencial de generar más de 38000 isoformas (Schmucker et al. 2000) Esto significa que el repertorio de proteínas que puede generar este gen es mayor que el número de genes de *D. melanogaster*.

Los intrones no sólo permiten pasivamente el proceso de corte y empalme alternativo por su mera existencia, sino que lo regulan activamente albergando elementos que regulan dicho proceso (Schwartz et al. 2008; Hartmann & Valcarcel 2009). Hay motivos *cis* que generalmente se unen a factores de corte y empalme que

favorecen o reprimen el montaje del espliceosoma a un sitio de unión cercano. Un ejemplo de ello es *Nova-1*, una proteína de unión a ARN específica de neurona que funciona principalmente en el cerebro y que regula el proceso de corte y empalme alternativo mediante su unión a motivos intrónicos (Dredge & Darnell 2003).

### **Funciones asociadas al intrón eliminado**

Una vez un intrón es eliminado, tras el proceso de corte y empalme, normalmente forma parte de un complejo denominado *post-splicing*, que conduce a la degradación del intrón de manera eficiente (Yoshimoto et al. 2009). Sin embargo cuando dentro del intrón se encuentra un gen ARN, éste se expresa una vez se ha producido la eliminación del intrón. Se han caracterizado varias familias de ARNs no codificantes, tales como, microARNs y pequeños ARNs nucleolares. Se han encontrado miembros de estas familias asociados a intrones en humanos, conduciendo a la hipótesis de que los genes pueden autorregular su expresión albergando ARNs no codificantes dentro de sus intrones (Chorev & Carmel 2012).

Una suposición fundamental en genómica comparativa es que una conservación evolutiva es indicativa de una función biológica. Esta conservación se identifica principalmente mediante una conservación en la secuencia de ADN, aunque también se detecta en la conservación de la estructura secundaria y terciaria del ADN, ARN y proteínas, además de una conservación en la organización del genoma (Graur & Li 2000). Los intrones invadieron en gran número a los primeros genomas eucariotas como elementos ligeramente sin importancia, pero más tarde ganaron muchas funciones hasta el punto de que hoy los eucariotas superiores no pueden sobrevivir sin ellos (Lynch 2007). Este factor implica que el nivel de conservación de un intrón y su posición puede correlacionarse con la importancia funcional de dicho intrón (Chorev & Carmel 2012).

## **9. FILOGENIA MOLECULAR BASADA EN EL GEN DEL CITOCROMO B MITOCONDRIAL**

La filogenia molecular sirve para determinar las relaciones evolutivas entre diferentes especies mediante la comparación de secuencias de ADN de genes ortólogos de cada una de ellas. Para llevar a cabo un buen estudio filogenético, es necesario conocer el ritmo y el modo de evolución de la secuencia elegida. A la hora de elegir una

secuencia determinada hay que tener en cuenta varios factores, tales como, que las secuencias codificantes están sujetas a un proceso de selección, lo que provoca que evolucionen más lentamente, ya que sólo acumulan las mutaciones que no afectan a la función de la proteína resultante, mientras que las secuencias no codificantes acumularían mutaciones más rápidamente. También hay que valorar los mecanismos de variación que afectan a cada gen. Los genes que presentan varios alelos y sufren recombinación, como algunos genes nucleares, y los que experimentan fenómenos de conversión génica o están sujetos a selección de balance, como es el caso de los genes del Sistema Principal de Histocompatibilidad, pueden hacer que el análisis filogenético de resultados poco fiables.

El gen que se suele usar con más frecuencia en los análisis filogenéticos es el gen del citocromo b mitocondrial. Se trata de un gen de 1,143 pares de bases que se localiza entre las posiciones 14,747 y 15,887 (Anderson et al. 1981), el cual produce una única proteína que está altamente conservada. Esta proteína posee entre ocho (Howell & Gilbert 1988) y nueve (Saraste 1984; Widger et al. 1984) dominios apolares (hidrofóbicos) con estructura de alfa hélice que atraviesan la membrana interna de la mitocondria formando dominios extramembranales. Contiene a su vez dos dominios hemo, denominados bk y br, cada uno de ellos ligado a los dominios transmembrana 2 y 4 mediante dos histidinas invariantes en las posiciones 96 y 182 (en el caso del dominio hemo bk) y 82 y 198 (para el dominio hemo br) (Barber 1984). Estas posiciones se ha visto que son invariantes en todas las especies estudiadas hasta el momento. La mayoría de las posiciones variables de esta proteína se localizan en los segmentos transmembrana, siendo la parte más externa de la proteína la que parece evolucionar más lentamente.

La utilidad del gen del citocromo b mitocondrial frente a otros genes en los estudios filogenéticos radica principalmente en su modo particular de evolución, el cual es muy conocido y estudiado, así como características propias del ADN mitocondrial, como son la ausencia de alelos y de recombinación.

Para realizar el análisis filogenético existen varios métodos, siendo unos de los más utilizados, aquellos que se basan en distancias genéticas. Esta distancia se refiere a la divergencia genética que existe entre especies diferentes o entre poblaciones de una misma especie. Si entre los individuos de estudio existe una relación genética estrecha la distancia genética obtenida será pequeña, mientras que una distancia genética grande indica una relación genética más distante.

Existen varios métodos que permiten el cálculo de las distancias genéticas y la construcción de árboles filogenéticos a partir de dichas distancias. Entre ellos se encuentran los descritos a continuación:

### ***9.1 Modelos de sustitución nucleotídica***

En estos modelos, la distancia genética existente entre dos secuencias de ADN se calcula observando las diferencias entre ellas. La forma más sencilla de calcular esta distancia es contar el número de posiciones variables a nivel de nucleótido existentes entre las dos secuencias y dividir este número entre el total de nucleótidos. Esta distancia se denomina “distancia p” (Nei 1972; Nei 1987). Esta distancia asume que los cambios de nucleótido ocurren con la misma frecuencia y en cualquier parte de la molécula, sin embargo el proceso evolutivo de un gen no es tan sencillo, ya que hay determinados tipos de mutaciones que ocurren con mayor frecuencia que otras y no siempre se distribuyen de manera homogénea a lo largo de toda la secuencia. Debido a esto, se han desarrollado varios modelos de sustitución nucleotídica que tienen en cuenta varias variables a la hora de calcular la distancia genética existente entre dos secuencias. Algunas de estas variables son:

**Frecuencias nucleotídicas:** se pueden dar los mismos valores de frecuencia a cada uno de los cuatro tipos de nucleótido o dar valores diferentes, lo que más se aproxime a la realidad.

**Tipos de sustituciones nucleotídicas:** se puede asumir que cualquier cambio de nucleótido es igual de probable o bien que cada tipo tiene una probabilidad diferente. El tipo de sustitución que se asume con más frecuencia es el que considera dos tipos de cambios de nucleótido, las transiciones y las transversiones. El primer tipo se basa en cambios entre nucleótidos del mismo tipo (púrica o pirimidínica) ( $A/G = C/T$ ). Mientras que las transversiones se refieren a cambios entre nucleótidos de distinto tipo ( $A/C = A/T = G/C = G/T$ ). Las transiciones suelen ser más frecuentes que las transversiones.

**Tasas de los tipos de sustitución:** los tipos de sustituciones se dan con una probabilidad o tasa que puede ser calculada de forma experimental, definida por el usuario o estimada como un valor de probabilidad. Los programas de filogenia molecular nos dan una tasa de sustitución por cada tipo de sustitución definido.

**Proporción de sitios invariantes:** el número de posiciones conservadas o variables entre un grupo de secuencias de ADN afecta al resultado del análisis

filogenético. De tal forma que, un número elevado de posiciones conservadas no daría un resultado filogenético fiable debido a la falta de elementos variables necesarios para realizar el análisis, y por el contrario un número elevado de elementos variables también afectaría al análisis ya que indicaría un ritmo de evolución elevado que puede estar enmascarando dobles cambios en una posición.

### ***9.2 Métodos de construcción de árboles filogenéticos***

Una vez obtenidas las distancias genéticas entre las secuencias de estudio de acuerdo con un modelo de sustitución nucleotídica, se puede mostrar el resultado de forma gráfica en forma de árbol filogenético o dendrograma. De esta manera se pueden reflejar las relaciones evolutivas que conectan a las secuencias consideradas en el estudio. Los métodos de construcción de árboles filogenéticos se describen a continuación:

***Método Neighbor-joining (NJ) (Saitou & Nei 1987):*** Este método parte de una matriz de secuencias de ADN y analiza todos los árboles resultantes de agrupar dos secuencias, dejando el resto desagrupadas. Una vez ha construido todos los posibles árboles (uno por cada posible pareja de secuencias) se calcula la longitud total de cada uno y se selecciona como válido el que tenga la menor longitud total. La longitud total del árbol es la suma de la longitud de cada una de sus ramas, siendo estas las distancias genéticas que separan los extremos de cada rama. Las secuencias quedan agrupadas de manera que se minimiza la distancia genética existente entre ellas tomadas en su conjunto.

***Método Maximum Likelihood (ML) (Felsenstein 1981):*** Este método evalúa la probabilidad de obtener un árbol a partir de una matriz dada, de acuerdo con los valores que toman las variables que definen el modelo de sustitución nucleotídica. El árbol que obtendremos será el que tenga mayor valor de probabilidad, denominado *likelihood score*. Los valores que toman las variables que definen el modelo de sustitución nucleotídica, pueden ser estimadas durante el proceso o bien pueden ser tomadas de un árbol NJ previamente construido. El árbol obtenido puede ser posteriormente linealizado de acuerdo al modelo de Thorne (Thorne et al. 1998), que admite un ritmo de cambio variable a lo largo del tiempo y distinto para cada linaje, con el fin de conocer la antigüedad relativa de cada grupo.

***Método de Inferencia Bayesiana (BI, Bayesian Inference) (Huelsenbeck & Ronquist 2001):*** Este método supone un avance frente a otros métodos similares, como

el *Maximum Likelihood* comentado anteriormente, ya que modifica los valores de las variables que definen el modelo de sustitución nucleotídica durante el proceso de construcción del árbol filogenético, de esta forma va creando árboles, que va mejorando en función de un valor de probabilidad. Para ello comienza construyendo un árbol a partir de una matriz de secuencias de ADN y teniendo en cuenta el modelo de sustitución nucleotídica programado por el usuario. A continuación calcula la probabilidad de ese árbol y a partir de este punto se repite el proceso inicial un número determinado de veces (generaciones). En cada generación se modifican los parámetros iniciales del modelo de sustitución nucleotídica con el fin de mejorar el valor de probabilidad y por tanto el resultado final. Gracias a este método obtenemos un árbol que va a tener en cada uno de sus nodos un valor de probabilidad o *bootstrap*, que nos va a indicar la confianza o veracidad de esas uniones. De esta forma se puede saber que nodos son más fiables y cuáles no. Cuanto más alto es el valor de *bootstrap*, mayor probabilidad hay de que esa unión sea un reflejo de la realidad. Mientras que valores bajos indicarían que esos nodos se habrían creado simplemente por azar.

### 10. LA CLASE AVES

La clase aves forma parte de las cinco clases de vertebrados que componen el filo Cordados del Reino Animal. Alberga casi 10,000 especies diferentes distribuidas en 23 órdenes y 2,068 géneros (Sibley 1995). Se trata de animales homeotermos, cuya característica principal que les distingue del resto de animales consiste en que su cuerpo está cubierto de plumas. Su esqueleto es resistente y liviano debido a la fusión de sus elementos y a la presencia de sacos aéreos. Además poseen extremidades delanteras modificadas que dan lugar a las alas que les proporciona la capacidad de volar, lo que ha permitido su extensión por todas las regiones del planeta.

Aunque el registro fósil de las aves es más reducido si lo comparamos con el de los peces o el de los reptiles, se han encontrado restos de los antecesores de las aves actuales de más de 150 millones de años, los cuales se remontan al Mesozoico. La historia evolutiva de las aves ha suscitado siempre mucho debate entre la comunidad científica. La mayoría de los paleontólogos coinciden en que las aves y los dinosaurios están muy relacionados, sin embargo el problema surge cuando se intenta explicar cómo es esa relación. Debido a esto se han generado dos teorías principalmente. En primer lugar hay científicos que proponen que las aves son una rama perteneciente al linaje de los dinosaurios que sobrevivieron a la crisis del Cretácico y que evolucionaron para dar

lugar a las aves actuales. La segunda teoría sugiere que hubo un ancestro común que dio lugar a ambos grupos. Por lo que según la última teoría, las aves no serían dinosaurios, pero si representarían el grupo de animales actuales más cercano a los dinosaurios extintos. La comunidad científica no se inclina hacia ninguna teoría firmemente debido a que existen evidencias a favor y en contra de ambas hipótesis relacionadas con la fisiología, la anatomía y la biología del desarrollo.

Aunque como se ha comentado anteriormente el registro fósil de las aves es bastante más escaso que el de otros animales, se han encontrado fósiles de lo que se cree es el fósil más antiguo que posee características similares a las de un ave. Este animal estaría en un estadio intermedio entre los reptiles y las aves modernas. Se trata del *Archaeopteryx* (Figura I8), una especie que vivió durante el Jurásico tardío. Los fósiles de esta especie se han conservado tan bien que hasta se pueden apreciar las plumas. Este animal poseía características similares a las aves actuales y a los Terópodos (grupo de dinosaurios propuesto como antepasado de las aves). Tenía grandes cuencas para los ojos, pico y tres dedos en las patas traseras, sin embargo, a diferencia de las aves actuales tenía unos dientes bien desarrollados y no tenía una quilla bien desarrollada, ya que su esternón era liso. Poseía una cola muy parecida a la de los dinosaurios pequeños, pero a diferencia de ellos tenía una hilera de plumas a cada lado. Todas estas características nos muestran un animal que no sería un experto volador, pero si sería capaz de dar saltos, planear y correr (Encyclopaedia Britannica 1999).

En lo que respecta a la capacidad de volar, para que se pudiera desarrollar, se tuvo que incrementar el tamaño de los músculos que posteriormente moverían las alas y desarrollarse una quilla en el esternón, para aumentar la superficie del mismo y así soportar los músculos utilizados en el vuelo. La cola por su parte, empezó a usarse como un timón, por lo que se tuvo que hacer más corta y ganar movilidad. Además el cerebro, el sistema circulatorio y el sistema respiratorio tuvieron que desarrollarse para mejorar la capacidad homeotérmica. En cuanto a las plumas, existe un debate abierto sobre la función que estas podían realizar en el momento de su aparición. Por un lado defiende su función reguladora de la temperatura corporal, ya que harían de aislante, y por otro lado se piensa que la función inicial de estos elementos estaría relacionada con la capacidad de volar (Encyclopaedia Britannica 1999).



### **10.1 Orden *Passeriformes***

Este orden es el más amplio y diverso dentro de la clase Aves. Actualmente de las casi 10,000 especies de aves existentes, más de la mitad ( $\approx 5,800$ ) son aves *Passeriformes*. Su distribución es mundial, alcanzando los niveles más altos de diversidad en los trópicos. Este orden se divide a su vez en dos subórdenes, *Passeri* y *Tyranni*. Las aves que forman parte del suborden *Passeri* son comúnmente llamadas pájaros cantores o aves de canto. Está formado por unas 4,000 especies y la característica principal que los diferencia de los *Tyranni*, es la de tener un aparato fonador complejo denominado siringe y su capacidad para aprender a cantar. La siringe se encuentra en una cavidad localizada en la zona de unión de la tráquea y los dos bronquios primarios. El sonido se produce cuando el aire es impulsado desde el saco de aire interclavicular por la contracción de los músculos torácicos y abdominales a través de la siringe provocando que los pasillos de ésta vibren. La complejidad de los músculos de la siringe es la que determina la calidad del sonido que ésta emite, así las aves que forman el suborden *Tyranni* (alrededor de 1,000) tendrían una musculatura simple, lo que provocaría que sólo pudieran emitir gruñidos o silbidos, mientras que las denominadas aves de canto (*Passeri*), tendrían una musculatura muy compleja como se ha mencionado anteriormente, lo que les daría la capacidad de producir sonidos más complejos (Greenewalt 1969; Gaunt & Wells 1973). Además hay que destacar que las aves pertenecientes al suborden *Tyranni* no tienen la capacidad de aprender sonidos nuevos, sino que el conocimiento de los mismos lo obtienen de manera innata.

### **10.2 El género *Serinus***

Este género comprende 45 especies (Sibley & Ahlquist 1990) generalmente de pequeño tamaño. Su hábitat se centra principalmente en África, pero también podemos encontrar algunas especies en Oriente Medio, Europa, Asia y los archipiélagos de la Macaronesia. La mayoría son aves pequeñas de color verdoso, limón, marrón o gris estriado (Clement et al. 1993). Se postuló que los ancestros de este género podían provenir de Eurasia (probablemente de Oriente Medio) (Clement et al. 1993), sin embargo estudios posteriores muestran como el origen podría haber estado en África, donde se habrían producido distintas radiaciones evolutivas (Zamora et al. 2006).

Según datos moleculares, la edad aproximada de aparición en la tierra de este género es de 9 millones de años. 2 millones de años tras su aparición se dividieron en dos grupos principalmente (Arnaiz-Villena et al. 2012). Uno de ellos dio lugar a los



canarios africanos, que a su vez se divide en tres grupos más, canarios africanos grandes (*S. sulphuratus*, *S. gularis*, *S. flaviventris* y *S. albogularis*), canarios africanos pequeños (*S. leucopygius*, *S. citrinelloides*, *S. dorsostratus*, *S. atrogularis*, *S. citrinipectus* y *S. mozambicus*) y canarios africanos basales (*S. striolatus*, *S. totta*, *S. olivaceus* y *S. thibetanus*). El segundo grupo principal incluiría a las especies que salieron de África y conquistaron otros hábitats en otros continentes. A su vez este segundo grupo se divide en canarios mediterráneos (*S. canaria* y *S. serinus*) y canarios afroasiáticos (*S. pusillus*, *S. syriacus*, *S. canicollis flavivertex*, *S. canicollis* y *S. alario*) (Arnaiz-Villena et al. 2012). El grupo más antiguo sería el de los canarios africanos, por lo que se asumiría un origen africano como se ha mencionado anteriormente (Zamora et al. 2006).

### ***10.3 Aves incluidas en el estudio***

Para este proyecto de tesis se han elegido 19 especies pertenecientes al género *Serinus*. Además se han añadido al estudio dos especies más, el pinzón vulgar (*Fringilla coelebs*) y el pollo (*Gallus gallus*). La descripción y distribución de cada una de ellas se detalla a continuación.

### *Serinus atrogularis* (canario de obispillo amarillo)

Se trata de un pájaro pequeño, que mide entre 11 y 12 centímetros de longitud y de color marrón grisáceo. Tiene un pico corto de forma cónica, un supercilio estrecho de color blanco, una mancha negra en la mejilla y otra en la garganta, rayas en los laterales del pecho y un obispillo de color amarillo. Los machos y las hembras son muy similares. Se distribuye por todo el sur de África, habitando desde sabanas hasta bosques y cultivos (Clement et al. 1993).



### *Serinus citrinelloides* (canario africano)

Esta especie suele medir entre 11, 5 y 12 centímetros de longitud. Su plumaje es brillante y variable entre las diferentes subespecies. Algunas presentan la cara negra con un supercilio amarillo. La parte superior es verde y suele tener un rayado negro. La parte inferior es del color amarillo brillante o amarillo verdoso apagado con rayas en el pecho y en los flancos. Se distribuye por Eritrea, Etiopía, Sudán, Uganda, Y Kenia principalmente (Clement et al. 1993).



### *Serinus citrinpectus* (canario de pecho limón)

Esta especie suele medir unos 12 centímetros de longitud. Los machos tienen el pecho y la barbilla de color amarillo limón, lo que les confiere su nombre común. Además tienen un punto amarillo muy brillante en la base de las mejillas. Su parte inferior es blanca. Las hembras y los juveniles son bastante similares, con un patrón facial más difuso. Ambos sexos tienen un obispillo amarillo brillante. Se distribuyen por Malawi, Zimbabue, Mozambique, Sudáfrica y Zambia (Clement et al. 1993).



### *Serinus dorsostriatus* (canario de vientre blanco)

Se trata de un pájaro de tamaño medio, oscilando entre 11,5 y 13 centímetros de longitud. La parte inferior del pecho y la parte superior del vientre son amarillas, mientras que la parte inferior del vientre y los flancos son blancos. Las hembras tienen un color más apagado que los machos, además de tener bandas en el pecho y en los flancos. Se distribuye por Kenia, Tanzania, Etiopía, Somalia y Uganda. Habita en matorrales áridos y semiáridos, en bosques, en áreas de sabana herbácea y en cultivos (Clement et al. 1993).



### *Serinus flaviventris* (canario amarillo)

Se trata de un canario de tamaño medio, con una longitud de 13 a 14 centímetros. Sus partes superiores son verdes o amarillas. Las hembras son más pálidas y suelen tener bandas. Se distribuye por Sudáfrica, Namibia, Botswana y Lesoto. Suele habitar matorrales y matorjos en zonas secas y semiáridas a los pies de las montañas (Clement et al. 1993).



### *Serinus gularis* (canario de cabeza estriada)

Se trata de un pájaro de unos 15 centímetros de longitud que posee un patrón facial y de cabeza bastante característico. Tiene un supercilio blanco y un píleo rayado. Sus mejillas son de color marrón oscuro, el mentón es blanco y el obispillo de color marrón. Ambos sexos son bastante similares. Se distribuye por zonas del centro y sur de África, como Botswana, Zimbabwe, Sudáfrica, Mozambique, Costa de Marfil y Camerún. Se encuentra en gran variedad de hábitats, que incluyen sabanas, jardines, bosques y huertas (Clement et al. 1993).



### *Serinus mozambicus* (canario de frente amarilla)

Esa especie mide entre 11 y 13 centímetros de longitud. Se trata de un pájaro de que posee un patrón facial muy marcado. Tiene la frente y el píleo de color verde mientras que el obispillo y las partes bajas son de color amarillo. Ambos sexos son muy similares aunque las hembras y los juveniles tienen un color ligeramente más apagado. Los juveniles además tienen rayas en las partes bajas. Se distribuye por el centro y sur de África, habitando las mesetas y bosques de sabana, los límites de los cultivos, jardines y alrededor de las casas (Clement et al. 1993).



### *Serinus striolatus* (canario estriado)

Este pájaro mide unos 15 centímetros de longitud. Tiene un pico bastante grande y es fácil de identificar gracias a su supercilio de color blanco, parte inferior rayada y alas y cola completamente marrones. Ambos sexos son muy similares. Se distribuye por Sudán, Etiopía, Kenia y Tanzania principalmente. Habita en bosques húmedos, matorrales, maleza, jardines y en los límites de los cultivos (Clement et al. 1993).





### *Serinus thibetanus* (canario tibetano)

Esta especie mide unos 12 centímetros de longitud, habiendo grandes diferencias entre ambos sexos. Los machos tienen el píleo y las partes superiores color verde oliva, mientras que el supercilio es de color amarillo. Las partes inferiores son verde amarillentas. Por su parte las hembras suelen ser más oscuras, con rayas en las partes superiores y en los flancos. El juvenil es similar a la hembra. Se distribuye desde el Himalaya hasta el sudeste del Tíbet. Puede verse también en la parte occidental de China. Habita laderas abiertas y bosques de coníferas, abedules, alisos, cicutas y abetos (Clement et al. 1993).



### *Serinus canaria* (canario silvestre)

Este pájaro que mide unos 12,5 centímetros de longitud. Es bastante similar al verdicillo y a otros miembros del género *Serinus* aunque es algo más grande y tiene una cola más larga. Su pico es de color rosáceo, su cara de un amarillo pálido y las partes inferiores son grises con rayas oscuras. Su obispillo es de color amarillo oliva y su cola oscura o gris. Las hembras suelen ser más oscuras, presentando rayas en los flancos. Se localiza en las Islas Azores, Madeira y Canarias, exceptuando las islas de Lanzarote y Fuerteventura. Suele estar en zonas elevadas, aunque también habita en jardines, huertos, bosques de laureles y pinos (Clement et al. 1993).



### *Serinus sulphuratus* (canario de Brimstone)

Se trata de un pájaro grande y robusto, que mide entre 15 y 16 centímetros de longitud. Tiene el supercilio de color amarillo y un pico bastante robusto. La parte superior suele ser verde con rayas negras, o amarillenta con un rayado oscuro, dependiendo de la región a la que pertenezca. En cuanto a la parte inferior, ésta suele ser amarilla completamente con algo de verde en los laterales del pecho. Ambos sexos son muy similares. Se localiza en el sur de África. Suele habitar llanuras costeras (Sudáfrica) y zonas con árboles dispersos, colinas y desfiladeros (Clement et al. 1993).



### *Serinus pusillus* (verdecillo carinegro)

Este canario mide entre 12,5 y 13 centímetros de longitud. El macho tiene la cabeza, la cara y la nuca negras a excepción de una mancha rojo brillante en la frente. El dorso es negro con bordes amarillos. El obispillo suele ser amarillo brillante o marrón claro. El pico es corto y achaparrado. La hembra sin embargo suele tener un color más pálido y en la frente presenta una fina banda rojiza. Su distribución abarca el centro y sur de Turquía, Irán, Afganistán, Pakistán, Asia central y el noroeste de China. Suele habitar en las laderas de montaña además de en prados alpinos (Clement et al. 1993).



### *Serinus canicollis* (canario de corona amarilla)

Se trata de un pájaro que mide unos 13 centímetros de longitud. Es de color verde apagado y se encuentra en el sur y el este de África. Los machos de las subespecies del sur tienen el píleo, nuca y los laterales del cuello de color gris pálido mientras que los machos de las subespecies del este tienen el píleo y la frente de color amarillo brillante. La nuca y el dorso son de color verde apagado. Las hembras también se pueden dividir según la cantidad de amarillo que presenten en la frente. Suele habitar en tierras bajas y áreas costeras (Clement et al. 1993).



### *Serinus serinus* (verdecillo)

Este pájaro mide 12 centímetros de longitud. Se trata de un pájaro pequeño con el pico corto y achaparrado. Los machos tienen la cabeza y el pecho de un color amarillo brillante. Las hembras son de color amarillo más pálido llegando a ser blanquecino en algunas ocasiones, además presentan un rayado negruzco. El obispillo de ambos es de color amarillo brillante. Su distribución abarca España, Francia, Turquía y Chipre. En general se encuentra en toda la costa mediterránea de África, Europa y Oriente próximo. Suele habitar jardines, parques, plantaciones y huertos (Clement et al. 1993).





### *Serinus leucopygius (canario de obispillo blanco)*

Se trata de un ave que mide unos 11 centímetros de longitud. Es un pájaro pequeño de color gris. La parte superior presenta manchas o rayas marrones y tiene un obispillo blanco muy característico. Se distribuye por el Sahara y el noroeste de África. Suele habitar zonas secas y abiertas de arbustos o en la sabana. También aparece en tierras cultivadas y jardines (Clement et al. 1993).



### *Serinus albogularis (canario de garganta blanca)*

Se trata de un pájaro bastante grande, ya que puede llegar a medir unos 15 centímetros de longitud. Es de color pardo, con la barbilla y la garganta de color blanco, un pico bastante prominente y un supercilio blanco bastante estrecho. Los 2 sexos son bastante similares. Se distribuye principalmente por Sudáfrica. Suele habitar en zonas de pasto seco, en las inmediaciones de los desiertos, en colinas, dunas, en los cursos de ríos secos y ocasionalmente en jardines (Clement et al. 1993).



### *Serinus alario (canario de cabeza negra)*

Esta ave mide entre 12 y 15 centímetros de longitud. Los machos tienen la cabeza, la nuca, el mentón, la garganta y la parte superior del pecho de color negro. La parte que va desde el dorso hasta el obispillo es de color castaño y las plumas utilizadas para el vuelo son de color marrón muy oscuro. Su distribución abarca Sudáfrica principalmente. Suele habitar en zonas secas: maleza, pastos, límites de cultivos y jardines suburbanos (Clement et al. 1993).



### *Serinus totta (canario de El Cabo)*

Este canario mide 13 centímetros de longitud. Las partes inferiores son de un verde amarillento muy apagado y la hembra posee una zona rayada que va desde el mentón hasta el pecho. El macho va a tener la cabeza y la cara de color verde oliva. La hembra, en cambio, tiene estas partes de color marrón y además presenta rayas oscuras. Su distribución se limita a la zona sudoeste de la provincia del El Cabo. Habita en montañas, colinas rocosas y acantilados costeros (Clement et al. 1993).



### *Serinus syriacus (verdecillo sirio)*

Se trata de un pájaro que mide entre 12 y 14 centímetros de longitud. Tiene una frente de color amarillo brillante. El mentón, la garganta y el área que rodea los ojos son de un amarillo más apagado. La espalda y las escapulares son de color verde amarillento. Su distribución se limita al Líbano, Siria e Israel. Habita en vertientes con arbustos, en montañas, en cedros y enebros y en otros árboles de hoja caduca (Clement et al. 1993).



### *Fringilla coelebs (pinzón vulgar)*

Este pájaro mide unos 14-18 centímetros de longitud, por lo que es más grande que las aves pertenecientes al género *Serinus* descritas anteriormente. Tiene un píleo ligeramente picudo y una cola ahorquillada. Tiene una forma de caminar característica, ya que va dando pequeños saltitos. Su distribución abarca casi toda Europa hasta Escandinavia, además de Rusia y Siberia. Suele habitar bosques de coníferas y de árboles de hoja caduca. También se le puede encontrar en huertos, jardines y parques (Clement et al. 1993).



### *Gallus gallus (pollo)*

Ave de gran tamaño, que puede llegar a medir hasta 50 centímetros de longitud. Esta especie dio origen al pollo doméstico actual. El macho tiene un plumaje característico, con tonalidades que van desde el rojizo al verde pasando por tonos negros y azulados. Tiene una cresta encima de la cabeza, los tarsos armados con un largo espolón y



barbillones que cuelgan a ambos lados del pico. La hembra es de tonalidad pardusca, más pequeña y no tiene ni cresta ni barbillones. Esta especie es originaria de Asia oriental, y solía habitar en bosques, selvas y zonas áridas, aunque en la actualidad se distribuye por todo el mundo debido a su gran valor comercial (Clement et al. 1993).



# OBJETIVOS



El estudio de los intrones está generando gran interés, ya que se ha visto que no están compuestos de “ADN basura”, sino que sus fragmentos de ADN juegan un papel importante en la expresión de muchos genes. El estudio de los intrones en aves silvestres de campo y su comparación con otras aves domésticas como es el caso del pollo, proporciona una visión natural evolutiva en aves, de estos fragmentos a lo largo de millones de años y puede arrojar datos importantes para la interpretación de la funcionalidad del ADN no codificante en el Sistema Principal de Histocompatibilidad.

Los objetivos que se plantean para este proyecto de tesis son:

- 1- Obtener las secuencias de ADN del intrón 2 de los genes MHC de clase I entre las aves silvestres de canto de un mismo género: el género *Serinus*.
- 2- Estudiar la evolución de la conservación del intrón 2 de los genes MHC de clase I mediante el análisis de las secuencias de ADN para determinar si hay posiciones variables, su localización y si estas posiciones se dan también en la secuencia del intrón 2 de otras aves silvestres y domésticas.
- 3- Estudiar la composición nucleotídica de la secuencia del intrón 2 de los genes MHC de clase I y compararla con la de otras aves.
- 4- Determinar si las aves de canto del género *Serinus* poseen un MHC “mínimo esencial” que implica que los intrones son inusualmente cortos, como es el caso del pollo.
- 5- Realizar un análisis evolutivo filogenético de especies hermanas basado en secuencias del gen del citocromo b mitocondrial, para ver su correlación con la evolución del intrón 2 de los genes MHC de clase I y comprobar si siguen un patrón de evolución similar.





# **MATERIALES Y MÉTODOS**



## 1. MUESTRAS EMPLEADAS

### 1.1 Muestras para el estudio del intrón 2 de los genes MHC de clase I

En este trabajo se han analizado muestras de 11 especies de aves pertenecientes al género *Serinus*. Los individuos utilizados fueron muestreados por toda la geografía mundial durante las últimas dos décadas. En la Tabla M1 se detallan las muestras empleadas y la procedencia de las mismas, así como el número de acceso a la base de datos del NCBI GenBank. Para completar el análisis de las secuencias del intrón 2 de los genes MHC de clase I, se añadieron a la comparación las secuencias del pinzón vulgar (*Fringilla coelebs*) y el pollo (*Gallus gallus*).

Tabla M1: Muestras utilizadas para el estudio del Intrón 2 de los genes MHC de clase I.

| Especie   | Origen de la muestra       | Número GenBank |
|---|----------------------------|----------------|
| <i>Serinus atrogularis</i><br>Canario de obispillo amarillo | Ciudad del Cabo, Sudáfrica | DQ257479       |
| <i>Serinus citrinelloides</i><br>Canario africano           | Nairobi, Kenia             | DQ257484       |
| <i>Serinus citrinipectus</i><br>Canario de pecho limón      | Maputo, Mozambique         | DQ257483       |
| <i>Serinus dorsostriatus</i><br>Canario de vientre blanco   | Dar es Salam, Tanzania     | DQ257486       |
| <i>Serinus flaviventris</i><br>Canario amarillo             | Ciudad del Cabo, Sudáfrica | DQ257487       |
| <i>Serinus gularis</i><br>Canario de cabeza estriada        | Ciudad del Cabo, Sudáfrica | DQ257489       |
| <i>Serinus mozambicus</i><br>Canario de frente amarilla     | Dar es Salam, Tanzania     | DQ257491       |
| <i>Serinus striolatus</i><br>Canario estriado               | Nairobi, Kenia             | DQ257493       |
| <i>Serinus thibetanus</i><br>Canario tibetano               | Szechwan, China            | DQ257496       |
| <i>Serinus canaria</i><br>Canario silvestre                 | Isla de La Palma, España   | DQ257480       |

|  |                            |          |
|--|----------------------------|----------|
| <i>Serinus sulphuratus</i><br>Canario de Brimstone | Ciudad del Cabo, Sudáfrica | DQ257494 |
| <i>Fringilla coelebs</i><br>Pinzón vulgar          | Madrid, España             | DQ257477 |
| <i>Gallus gallus</i><br>Pollo                      | Brasil                     | AY489160 |

### ***1.2 Muestras para el estudio filogenético basado en el gen del citocromo b mitocondrial.***

Para interpretar el proceso evolutivo al que han estado sometidas las especies de estudio y así conocer su relación filogenética, se recurrió al gen del citocromo b mitocondrial. En la Tabla M2 se pueden observar de forma detallada las muestras empleadas para este análisis, así como su procedencia y número de acceso de GenBank. Para este análisis se ha aumentado el número de especies hasta 20 buscando que los árboles filogenéticos obtenidos sean lo más precisos posible. A mayor número de especies del género *Serinus*, más similares a la realidad serán las relaciones filogenéticas obtenidas.

Tabla M2: Muestras utilizadas para el estudio filogenético basado en el gen del citocromo b mitocondrial.

| <b>Especie</b>  | <b>Origen de la muestra</b> | <b>Número GenBank</b> |
|---|-----------------------------|-----------------------|
| <i>Serinus atrogularis</i><br>Canario de obispillo amarillo | Ciudad del Cabo, Sudáfrica  | L76267                |
| <i>Serinus citrinelloides</i><br>Canario africano           | Nairobi, Kenia              | L77555                |
| <i>Serinus citrinipectus</i><br>Canario de pecho limón      | Maputo, Mozambique          | L78707                |
| <i>Serinus dorsostriatus</i><br>Canario de vientre blanco   | Dar es Salam, Tanzania      | L76278                |
| <i>Serinus flaviventris</i><br>Canario amarillo             | Ciudad del Cabo, Sudáfrica  | L76280                |
| <i>Serinus gularis</i><br>Canario de cabeza estriada        | Ciudad del Cabo, Sudáfrica  | L77556                |
| <i>Serinus mozambicus</i><br>Canario de frente amarilla     | Dar es Salam, Tanzania      | L76265                |

|   |                            |          |
|---|----------------------------|----------|
| <i>Serinus striolatus</i><br>Canario estriado             | Nairobi, Kenia             | L77557   |
| <i>Serinus thibetanus</i><br>Canario tibetano             | Szechwan, China            | L76279   |
| <i>Serinus canaria</i><br>Canario silvestre               | Isla de La Palma, España   | L76266   |
| <i>Serinus sulphuratus</i><br>Canario de Brimstone        | Ciudad del Cabo, Sudáfrica | L76294   |
| <i>Serinus pusillus</i><br>Verdecillo carinegro           | SinWiang, China            | L77873   |
| <i>Serinus</i><br>Verdecillo                              | Madrid, España             | L76263   |
| <i>Serinus canicollis</i><br>Canario de corona amarilla   | Ciudad del Cabo, Sudáfrica | L78706   |
| <i>Serinus leucopygius</i><br>Canario de obispillo blanco | Dakar, Senegal             | L76264   |
| <i>Serinus albogularis</i><br>Canario de garganta blanca  | Ciudad del Cabo, Sudáfrica | L78705   |
| <i>Serinus alario</i><br>Canario de cabeza negra          | Ciudad del Cabo, Sudáfrica | L76276   |
| <i>Serinus totta</i><br>Canario de El Cabo                | Ciudad del Cabo, Sudáfrica | AY570548 |
| <i>Serinus syriacus</i><br>Verdecillo sirio               | Mont. Hormon, Israel       | AY570547 |
| <i>Fringilla coelebs</i><br>Pinzón vulgar                 | Madrid, España             | L76609   |

## 2. OBTENCIÓN DEL MATERIAL GENÉTICO

Todas las muestras utilizadas en este estudio pertenecen a individuos salvajes capturados en sus entornos naturales y han sido obtenidas por el Dr. Arnaiz Villena a lo largo de los últimos 20 años. El proceso de obtención de las muestras ha sido el siguiente:

En primer lugar, se captura al ave mediante reclamo y unas redes muy finas; Se obtienen unas tres gotas de sangre mediante el corte de una uña previa anestesia de la zona con crema de lidocaína (EMLA, Laboratorios Astra, Suecia); Las gotas de sangre se depositan en tubos estériles de 5 ml junto con una solución conservante formada por

cloruro sódico 5M y dimetilsulfóxido (DMSO) en una proporción de 3:1. Finalmente los tubos se conservan a 4°C hasta su posterior almacenamiento a -20°C.

### 2.1 Extracción de ADN

A partir de las muestras de sangre obtenidas, se extrajo el ADN de los eritrocitos, los cuales son nucleados en el caso de las aves, lo que facilita el proceso de extracción, ya que con pocas gotas de sangre podemos obtener suficiente material genético para posteriores análisis. El proceso de extracción del ADN se llevó a cabo mediante el uso de un extractor automático (*Nucleic Acid Extraction System, QuickGene-810, FUJIFILM, Japón*) y un kit comercial indicado para la extracción de ADN a partir de sangre total (*QuickGene Whole Blood Extraction Kit S, FUJIFILM, Japón*). El proceso de extracción es el siguiente:

En primer lugar se prepara la mezcla que posteriormente se pasará por el extractor automático, para ello hay que seguir los siguientes pasos:

- Diluir 50 µl de sangre en 150 µl de solución salina (PBS pH 7.4, GIBCO, INVITROGEN, Auckland, USA) en un tubo de 1.5 ml.
- Añadir 30 µl de Proteasa (*DNA whole blood Protease EDB-01, FUJIFILM, Japón*) y 250 µl de solución de lisis (*DNA whole blood Lysis Buffer LDB-04, FUJIFILM, Japón*). Mezclar bien mediante vortex.
- Incubar la mezcla obtenida en un baño a 56°C durante 10 minutos.
- Añadir 250 µl de etanol absoluto (*Ethanol Absolute PRS – CH<sub>3</sub>CH<sub>2</sub>OH, PANREAC, España*). Mezclar bien mediante vortex.

En algunos casos donde la cantidad de ADN que hay en la muestra de sangre es elevada, al añadir el etanol absoluto el ADN precipita formándose lo que se denomina “medusa”. En estos casos hay que retirar la medusa antes de introducir la muestra en el extractor automático, ya que de no hacerlo, quedaría retenida en la membrana de la columna que se introduce en el extractor y se perdería. Esta medusa hay que procesarla de la siguiente forma:

- Añadir 600 µl de etanol absoluto (*Ethanol Absolute PRS – CH<sub>3</sub>CH<sub>2</sub>OH, PANREAC, España*) y dejar evaporar.

- Añadir 600 µl de etanol al 70% (*Ethanol 70% v/v BP CODEX – CH<sub>3</sub>CH<sub>2</sub>OH*, PANREAC, España). (Repetir este paso hasta que la medusa este blanquecina).
- Resuspendemos en 200 µl de agua bidestilada o en la solución de elución (*DNA whole blood Elution Buffer CDB-02, FUJIFILM, Japón*) proporcionada por el kit de extracción.

El sobrenadante que nos ha quedado al retirar la medusa (en el caso de que se haya formado) se procesa a través del extractor automático (*Nucleic Acid Extraction System, QuickGene-810, FUJIFILM, Japón*). Los pasos son los siguientes:

- Pasar la muestra a la columna del kit. Estas columnas poseen una membrana de 80 µm de espesor donde quedan atrapadas las moléculas de ADN, además son hidrófilas por lo que no se fijan a ella ni proteínas ni lípidos.
- Colocar la columna en el extractor automático junto con el tubo de recepción de ADN y el recipiente de descarte de lavado los cuales proporciona el kit.
- Proceder a la extracción del ADN. Esto se lleva a cabo mediante tres ciclos de lavado (*DNA whole blood Wash Buffer WDB-03, FUJIFILM, Japón*) y la posterior dilución en 200 µl de agua bidestilada o en la solución de elución (*DNA whole blood Elution Buffer CDB-02, FUJIFILM, Japón*).
- Una vez terminado el proceso de extracción se obtiene una cantidad de ADN de entre 150 y 200 µl.

Para conocer la concentración y pureza del ADN obtenido se midió la absorbancia mediante un espectrofotómetro (*ND-1000 spectrophotometer, NanoDrop, DE, USA*), utilizando una cantidad de ADN de 1.5 µl. El espectrofotómetro mide la absorbancia de la muestra a diferentes longitudes de onda: 240, 260 y 280 nm para así calcular su concentración. Recordar que una unidad de absorbancia a 260 nm corresponde a una concentración de 50 ng/µl. También se puede conocer la pureza del ADN extraído, calculando el cociente entre las absorbancias a 260 y 280 nm ( $A_{260}/A_{280}$ ), la cual tiene que dar valores comprendidos entre 1.6 y 1.8.

De cada una de las muestras empleadas en este trabajo, se obtuvieron concentraciones que estaban comprendidas entre 100 y 200 ng/µl, diluyendo con agua



bidestillada o concentrando mediante evaporación en estufa a 60°C en los casos donde fuera necesario.

### 3. OBTENCIÓN DE SECUENCIAS DE ADN DEL INTRÓN 2 DE LOS GENES MHC DE CLASE I.

El sistema Principal de Histocompatibilidad se caracteriza por estar constituido por varios genes y por su alto polimorfismo. Para estudiar la variabilidad de este sistema en los individuos de estudio del género *Serinus*, se recurrió a las técnicas de amplificación, secuenciación y clonación.

#### *3.1 Amplificación del intrón 2 de los genes de MHC de clase I*

Para obtener una mayor cantidad de ADN y así poder realizar más análisis de las secuencias correspondientes al intrón 2 del MHC de clase I de las aves de estudio, se recurrió a la reacción en cadena de la polimerasa, conocida como PCR (del inglés: *Polymerase Chain Reaction*, (Saiki et al. 1985; Mullis & Faloona 1987)), ya que es una técnica que permite conseguir un número de copias de un fragmento de ADN muy superior al que se obtendría sólo con el proceso de extracción. Esta técnica está basada en los mecanismos naturales que usan las células para replicar el ADN. Para realizarla se necesita:

- El fragmento de ADN que se desea amplificar.
- Cebadores: fragmentos cortos de ADN que van a ser complementarios a los extremos 3' de la secuencia que se quiere amplificar.
- Los 4 tipos de nucleótidos libres (A, C, T y G).
- La enzima ADN polimerasa, la cual crea una secuencia nueva a partir de los nucleótidos libres y la secuencia de ADN que queremos amplificar.

La técnica de la PCR se desarrolla en tres pasos:

- **Desnaturalización:** Proceso en el que se separan las 2 hebras de ADN en dos cadenas. Se consigue calentando la muestra a la temperatura de 93-97°C. Esta temperatura depende de la proporción de G+C y de la longitud de la cadena de ADN.
- **Anillamiento o Hibridación:** Etapa de unión de los cebadores a la secuencia complementaria del ADN molde. Este proceso se da a una temperatura que

oscila entre 50°C y 65°C, dependiendo de la composición nucleotídica del ADN molde y especificidad de los cebadores utilizados.

- **Elongación:** Fase en la que la ADN polimerasa sintetiza la cadena complementaria al molde de ADN, utilizando los cebadores como soporte inicial. La temperatura en esta parte del proceso suele ser de 72°C, con el fin de que la ADN polimerasa alcance su rendimiento óptimo.

Cuando finaliza el ciclo de las tres etapas, se repite todo el proceso en un nuevo ciclo. Es importante resaltar que con cada ciclo se dobla el número de copias del ADN molde.

Como se ha mencionado anteriormente, se recurrió a la PCR para amplificar las secuencias correspondientes al intrón 2 de los genes de MHC de clase I. Las proteínas de histocompatibilidad de clase I son muy polimórficas registrando sus puntos más variables en los dominios extracelulares alfa 1 y alfa 2, que son los encargados de albergar al péptido microbiano para su posterior presentación a los linfocitos T. Los dominios alfa 1 y alfa 2 están codificados por los exones 2 y 3 respectivamente. Entre los cuales se encuentra el intrón 2, el cual es el objeto de estudio de este proyecto de tesis. En la Figura M1 se puede ver un esquema del fragmento del gen de MHC de clase I que se ha amplificado para cada una de las especies de aves estudiadas. Para obtener el intrón 2 al completo, se amplificó parte del exón 2 y casi todo el exón 3.

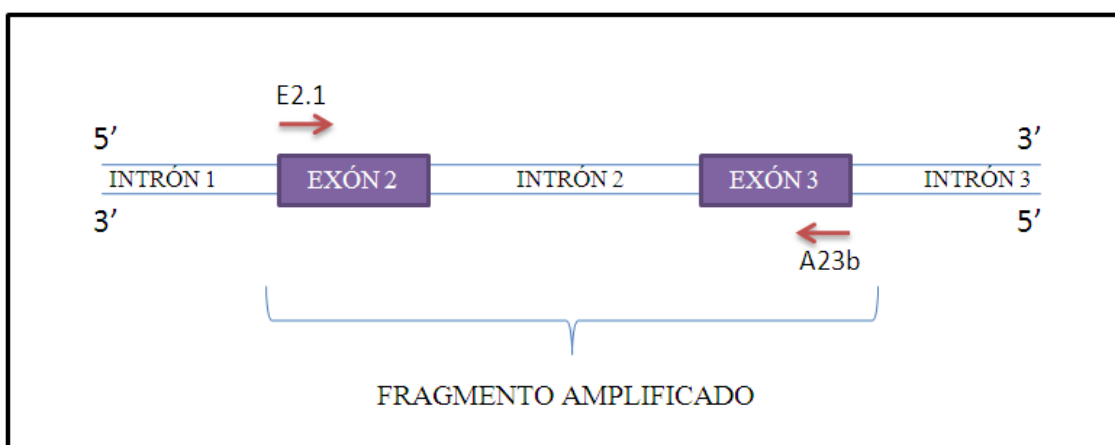


Figura M1: Esquema del fragmento del gen MHC de clase I amplificado.

Las reacciones de amplificación se llevaron a cabo en un termociclador (*Mastercycler epgradient S, Eppendorf, Alemania*). La ADN polimerasa utilizada fue

Taq DNA Polymerase (*Invitrogen, USA*), suministrada junto con el tampón (*10x PCR Buffer, Minus Mg<sup>++</sup>, Invitrogen, USA*), cloruro de magnesio (*50mM Magnesium Chloride, Invitrogen, USA*) y dNTPs (*100 mM dNTP Set, Invitrogen, USA*).

Los cebadores utilizados fueron: (Westerdahl et al. 1999).

- **Cebador Directo (E2.1):** 5'-GTTCTCCACTCCCTGGATTACC-3'
- **Cebador Inverso (A23b):** 5'- GCGCTCCAGCTCCTTCTGCCC<sup>A</sup>/GTA-3'

Las condiciones de amplificación y las características del programa utilizado se pueden ver en las Tablas M3 y M4 respectivamente.

Tabla M3: Condiciones del proceso de amplificación.

| Reactivos         | Concentración        | Volumen |
|-------------------|----------------------|---------|
| dNTPs             | 25 mM                | 1.6 µl  |
| Buffer            | 10x                  | 2 µl    |
| MgCl <sub>2</sub> | 2,5 mM               | 2 µl    |
| Cebador directo   | 10 µM                | 0,8 µl  |
| Cebador inverso   | 10 µM                | 0,8 µl  |
| Taq Polimerasa    | 5 U/µl               | 0,3 µl  |
| ADN               | 100-200 ng/µl        | 1 µl    |
| Agua              |                      | 11,5 µl |
|                   | <b>Volumen total</b> | 20 µl   |

Tabla M4: Características del programa de amplificación.

| Número de ciclos | Temperatura | Tiempo      | Proceso                   |
|------------------|-------------|-------------|---------------------------|
| 1                | 95°C        | 5 minutos   | Desnaturalización inicial |
|                  | 95°C        | 10 segundos | Desnaturalización         |
| 42               | 65°C        | 30 segundos | Hibridación               |
|                  | 72°C        | 60 segundos | Elongación                |
| 1                | 72°C        | 10 minutos  | Elongación final          |

### ***3.2 Control de la amplificación mediante electroforesis en gel de agarosa***

La técnica de PCR es un procedimiento muy sensible a contaminaciones, lo que puede dar lugar a amplificaciones no deseadas. Por ello hay que introducir un control negativo que tendrá mismos reactivos que el resto de las muestras a excepción del ADN, que se sustituirá por agua. Además de añadir un control negativo, tras la PCR, se realizó un control de la amplificación mediante electroforesis en gel de agarosa al 2%. Esta electroforesis permitió identificar y aislar los fragmentos correspondientes al gen de MHC de clase I. En este proceso se utilizó Agarosa (*Biotoools HR Agarose, High resolution, BIOTOOLS, España*) al 2% en un tampón 1X TBE (*10X TBE Buffer, Invitrogen, UK*). Se incluyó un marcador de peso molecular (*DNA Molecular Weight Marker XIV – 100 bp ladder, ROCHE, Alemania*) para conocer el tamaño del fragmento amplificado, de unas 850 pb. Para poder ver la presencia del fragmento de ADN deseado, se tiñó el gel con un agente intercalante (*Olerup SSP GelRed Dropper Bottle, Suecia*) y posteriormente se observó con luz ultravioleta. La banda correspondiente con el gen de MHC de clase I, de unas 850 pb fue recortada con un bisturí estéril y depositada en un tubo de 1.5 ml para su posterior purificación.

### ***3.3 Purificación de los productos de PCR***

Finalizado el proceso de electroforesis y comprobado que no hay contaminación en el control negativo se llevó a cabo la purificación del ADN obtenido. Para ello se utilizó un kit comercial de purificación (*MinElute Gel Extraction Kit (250), QIAGEN, Alemania*). Este kit está compuesto por unas columnas que poseen una membrana de sílice en la que se va a quedar pegado el ADN, mientras que los restos de gel, sales y dNTPs van a atravesar la membrana y ser eliminados a un tubo colector tras una serie de lavados.

El proceso de purificación es el siguiente:

- Añadir 300 µl del Buffer QG (*Solubilization Buffer QIAGEN, Alemania*) al tubo donde se encuentra la banda de gel recortada e introducirlo en un baño a 56°C hasta que el gel se disuelva.
- Pasar el contenido disuelto del tubo por la columna del kit.
- Centrifugar 1 minuto a 14000 rpm y decantar el filtrado.
- Añadir 500 µl del Buffer QG (*Solubilization Buffer QIAGEN, Alemania*).

- Centrifugar 1 minuto a 14000 rpm y decantar el filtrado.
- Añadir 650 µl del Buffer PE (*Wash Buffer QIAGEN, Alemania*) y dejar actuar durante 5 minutos.
- Centrifugar 1 minuto a 14000 rpm y decantar el filtrado.
- Centrifugar de nuevo 1 minuto a 14000 rpm y decantar el filtrado.
- Colocar la columna en un tubo de 1.5 ml.
- Añadir 50 µl del Buffer de Elución o de agua bidestilada.
- Centrifugar 1 minuto a 14000 rpm.
- Desechar la columna y conservar el ADN purificado a -20°C hasta su uso.

### ***3.4 Secuenciación de los productos purificados***

El proceso de secuenciación se basó en el modelo descrito por Sanger y colaboradores en los años 70 (Sanger et al. 1977). En el proceso de PCR de secuenciación, el ADN que se desea secuenciar va a funcionar como molde para la síntesis enzimática de un nuevo fragmento de ADN. Este proceso comienza en un punto definido por un cebador específico. Se usan desoxinucleótidos trifosfato (dNTPs) y didesoxinucleótidos trifosfato (ddNTPs); estos últimos al incorporarse en el proceso de síntesis del ADN, bloquean la elongación de la cadena, dando como resultado una población de fragmentos de ADN de diferente longitud. Para conocer que nucleótido termina la cadena en cada posición se suele realizar una reacción de elongación combinando los 4 tipos de didesoxinucleótidos marcando cada uno de ellos con un fluoróforo diferente. Las moléculas resultantes se separan realizando una electroforesis en gel de acrilamida y la secuencia se obtiene gracias a la relación entre el orden de los fragmentos en la electroforesis y el didesoxinucleótido en el que termina cada uno de ellos (Figura M2).

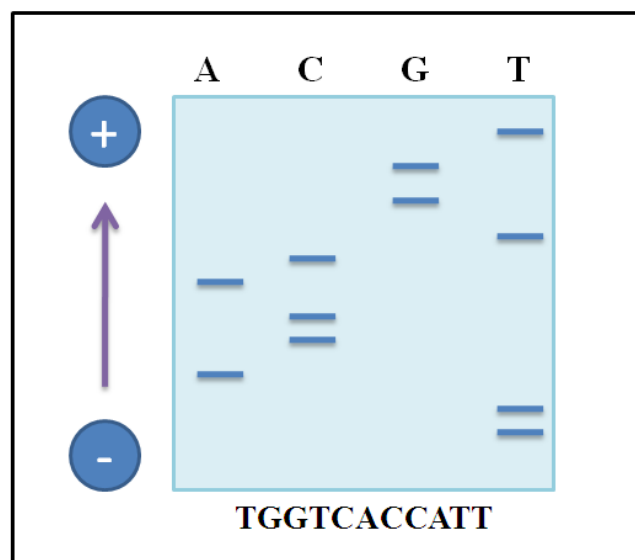


Figura M2: Ejemplo secuenciación método Sanger

Para llevar a cabo este proceso de secuenciación, se enviaron las muestras a la empresa SECUGEN S.L. ([www.secugen.es](http://www.secugen.es)), situada en el Centro de Investigaciones Biológicas (CIB-CSIC) de Madrid. Antes de ser enviadas, se prepararon las muestras utilizando 15 µl del ADN amplificado y purificado con una concentración de alrededor de 10 ng/ml, y 1.5 µl del mismo cebador que se usó en la amplificación, pero con una concentración de 5 mM.

### 3.5 Clonación de los productos purificados

La calidad de las secuencias recibidas tras el proceso de secuenciación, depende de la pureza y correcta cuantificación de las muestras de ADN analizadas. En ocasiones al analizar dichas secuencias encontramos ambigüedades que nos ofrecen dos o más alternativas para una misma posición nucleotídica como se puede ver en la Figura M3. Estas ambigüedades pueden indicar que el individuo del cual procede la muestra es heterocigoto para el gen amplificado o que se han amplificado dos loci simultáneamente con secuencias muy similares.

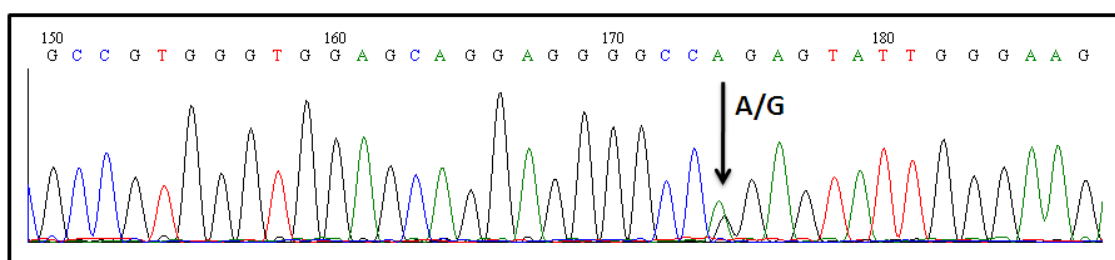


Figura M3: Secuencia de ADN con una ambigüedad en forma de doble pico.

Para resolver las posibles ambigüedades y saber si los cambios observados son reales y no se deben a errores durante el proceso de secuenciación, se recurrió al proceso de clonación.

La clonación se basa en la capacidad natural que tienen algunas bacterias para incorporar pequeñas moléculas circulares de ADN, denominadas plásmidos, presentes en su entorno, mediante un proceso denominado “transformación bacteriana” (Sambrook et al. 1989). Los plásmidos poseen genes que aportan ciertas ventajas o mejoras a las bacterias, como puede ser la resistencia a ciertos antibióticos. Estos son replicados en el interior de la bacteria y transmitidos a su descendencia. La eficacia de la transformación suele ser baja, y en general, la bacteria que la realiza con éxito incorpora un único plásmido.

En la actualidad existen plásmidos artificiales también denominados vectores, en los cuales se puede insertar material genético mediante técnicas de ingeniería genética. Gracias a esta tecnología de clonación, se pueden aislar y caracterizar secuencias de ADN de forma individual.

Para el proceso de clonación se utilizó el sistema *pGEM-T Easy* (PROMEGA). Este vector posee dos marcadores, el primero es un gen de resistencia a ampicilina, que nos permite seleccionar a las bacterias que han incorporado el plásmido, con o sin el fragmento de ADN deseado. El otro marcador es el lugar de inserción del fragmento que se encuentra en mitad del gen *lacZ*. Este gen codifica para la enzima  $\beta$ -galactosidasa, la cual metaboliza el X-gal y genera un compuesto azul insoluble. Si en la placa de cultivo se disuelve X-gal y un inductor de la  $\beta$ -galactosidasa (IPTG), las colonias de bacterias que tengan el gen *lacZ* funcional, es decir, que no posean el inserto, se tornaran de color azul, mientras que las que si posean el fragmento de ADN serán blancas, debido a que tendrán el gen *lacZ* truncado. De esta manera podemos distinguir aquellas bacterias que han transformado el vector con el inserto deseado.

El proceso de clonación es el siguiente:

- Preparar la reacción de ligado mezclando 5  $\mu$ l de tampón (*2x Rapid Ligation Buffer*, PROMEGA), 1  $\mu$ l de plásmido (*pGEM-T Easy Vector*, PROMEGA), 3  $\mu$ l de producto de amplificación y 1  $\mu$ l de ligasa (*T4 DNA Ligase*, PROMEGA).

- Incubar 90 minutos a temperatura ambiente o bien durante 12-16 horas a 4°C.
- Añadir 50 µl de células competentes (*JM109 Competent Cells, High efficiency*, PROMEGA) a un tubo conteniendo 2 µl de la mezcla de ligado.
- Incubar en hielo durante 20 minutos.
- Incubar en un baño a 42°C durante 40-45 segundos.
- Incubar en hielo durante 2 minutos.
- Añadir 950 µl de Medio SOC, que contiene 30.7 gr/l de Medio SOB (*SOB MEDIA*, PRONADISA) y Glucosa 20 mM (*DEXTROSA*, PRONADISA).
- Incubar a 37°C en agitación (150 rpm) durante 90 minutos.
- Sembrar 100 µl en placas con Medio LB Agar conteniendo ampicilina, IPTG y XGAL (2 placas por muestra). 1 litro de medio contiene 35 gr de LB Agar (*LB AGAR LENNOX*, PRONADISA), 10 ml de ampicilina 10 mg/ml (*Ampicillin Sodium SALT – Irradiated*, GIBCO), 0.5 ml de IPTG 1M (*IPTG Dioxane-Free*, PROMEGA) y 1.6 ml de X-Gal 50 mg/ml (*X-GAL*, PROMEGA).
- Incubar a 37°C durante 16-24 horas.
- Opcionalmente, las placas se pueden mantener a 4°C durante unas horas después de la incubación para favorecer la aparición de color azul en las colonias “negativas”.

### ***3.6 Obtención de las secuencias de ADN de los fragmentos clonados***

Tras el proceso de clonación se seleccionaron 20 colonias blancas por muestra para proceder a la extracción del ADN bacteriano mediante el siguiente procedimiento:

- Resuspender la colonia en 20 µl de agua destilada en un tubo de 0.2 ml.
- Incubar durante 5 minutos a 95°C en un termociclador.
- Incubar durante 5 minutos en hielo.
- Centrifugar 3 minutos a 12000 rpm.
- Pasar el sobrenadante (aproximadamente 15 µl) a un tubo de 1.5 ml. Conservar a -20°C.

Para finalizar se amplificaron las secuencias de los genes de MHC de clase I utilizando como ADN molde el extraído de las bacterias (el cual contenía el ADN bacteriano y el plásmido con el inserto). La PCR se realizó en las mismas condiciones en las que se amplificó el ADN genómico (ver apartado 3.1). Los productos



amplificados fueron sometidos a una electroforesis en gel de agarosa al 2% para separar el fragmento de ADN deseado (ver apartado 3.2), posteriormente fueron purificados (ver apartado 3.3) y finalmente se secuenciaron (ver apartado 3.4).

#### **4. ANÁLISIS DE LAS SECUENCIAS DE ADN DEL INTRÓN 2 DE LOS GENES MHC DE CLASE I.**

En este trabajo se han estudiado 11 individuos pertenecientes al género *Serinus*. Cada muestra de ADN de cada individuo ha sido sometida a un proceso de amplificación, secuenciación y clonación.

##### ***4.1 Caracterización de las secuencias de estudio.***

Para poder analizar la evolución de las secuencias de ADN del intrón 2 de los genes MHC de clase I, es necesario realizar en primer lugar un alineamiento múltiple entre las diferentes secuencias obtenidas. De esta manera se puede ver si existe homología a nivel de nucleótido. Como el intrón 2 es eliminado por el espliceosoma y por lo tanto no se traduce a proteína, no se realizó la comparación a nivel de secuencia proteica. Para realizar el alineamiento de las secuencias, se utilizó el programa MEGA 7 (*Molecular Evolutionary Genetics Analysis*) (Kumar et al. 2016). Las secuencias del intrón 2 de los genes MHC de clase I pertenecientes al género *Serinus* se compararon entre sí, además de con la secuencia de *Fringilla coelebs* y *gallus gallus*.

Los animales vertebrados de sangre caliente poseen regiones de hasta 300 kb ricas en nucleótidos de guanina y citosina (G+C). Se ha visto que estas regiones son más abundantes en los genomas de las aves que en mamíferos (Bernardi et al. 1997). Las regiones ricas en G+C son más estables frente a la desnaturalización térmica, lo que representa una ventaja para las aves, ya que tienen temperaturas y tasas metabólicas más elevadas (Bernardi & Bernardi 1986). En el género *Serinus*, grupo al que pertenecen nuestras aves de estudio, se ha observado que en el exón 2 de los genes de MHC de clase I el contenido de G+C supera el 60% (Lowy 2003). Para ver si esta característica se observa también en las secuencias del intrón 2 de nuestras aves de estudio, se realizó el cálculo de las frecuencias nucleotídicas del intrón 2 de los genes MHC de clase I. Para ello se utilizó el programa MEGA 7 (*Molecular Evolutionary Genetics Analysis*) (Kumar et al. 2016). Esta función permite calcular las frecuencias de cada nucleótido en una región deseada de una secuencia de ADN (en nuestro caso del intrón 2).

### ***4.2 Análisis del grado de conservación de las secuencias de estudio.***

Para estudiar si existe o no conservación en las secuencias del intrón 2 de los genes de MHC de clase I de las especies de estudio pertenecientes al género *Serinus*, se recurrió al cálculo de las distancias genéticas y al cálculo del porcentaje de similitud.

La distancia genética entre dos secuencias de ADN se obtiene tomando el número de posiciones variables existentes entre dos secuencias y dividiéndolo entre el número total de nucleótidos de la secuencia. A esta distancia se la denomina distancia p (Nei 1987) y se calcula entre pares de secuencias. Para el cálculo de las distancias genéticas se utilizó el programa MEGA 7 (Molecular Evolutionary Genetics Analysis) (Kumar et al. 2016). El porcentaje de similitud se calculó con el programa BioEdit (Hall 1999).

## **5. FILOGENIA DEL GÉNERO *SERINUS* BASADA EN EL GEN DEL CITOCROMO B MITOCONDRIAL**

Para conocer el proceso evolutivo al que han estado sometidas las secuencias del intrón 2 de los genes de MHC de clase I de las aves de estudio, es necesario estudiar las relaciones filogenéticas existentes entre ellas. Como los genes de Histocompatibilidad no están recomendados para analizar relaciones filogenéticas, debido a que tienen una forma de evolucionar un poco peculiar, se recurrió al gen del citocromo b mitocondrial.

Las secuencias utilizadas forman parte del archivo del grupo que dirige el Dr. Arnaiz Villena (ver Tabla M2). Estas secuencias ya fueron comparadas en trabajos anteriores (Arnaiz-Villena et al. 1999; Arnaiz-Villena et al. 2007). Para verificar los resultados previos, se han realizado una serie de análisis que se comentan a continuación.

### ***5.1 Construcción de un árbol filogenético basado en el método Neighbor-joining (NJ) (Saitou & Nei 1987)***

Este método parte de una matriz de secuencias de ADN y analiza todos los árboles resultantes de agrupar dos secuencias, dejando el resto desagrupadas. Una vez ha construido todos los posibles árboles (uno por cada posible pareja de secuencias) se calcula la longitud total de cada uno y se selecciona como válido el que tenga la menor longitud total. La longitud total del árbol es la suma de la longitud de cada una de sus

ramas, siendo estas las distancias genéticas que separan los extremos de cada rama. Las secuencias quedan agrupadas de manera que se minimiza la distancia genética existente entre ellas tomadas en su conjunto.

Para construir el árbol filogenético basado en este método se utilizó el programa MEGA 7 (Molecular Evolutionary Genetics Analysis) (Kumar et al. 2016).

### ***5.2 Construcción de un árbol filogenético basado en el método Maximum Likelihood (ML)***

El método ML (Felsenstein 1981) evalúa la probabilidad de obtener un árbol a partir de una matriz dada, de acuerdo con los valores que toman las variables que definen el modelo de sustitución nucleotídica. El árbol que obtendremos será el que tenga mayor valor de probabilidad, denominado *likelihood score*. Los valores que toman las variables que definen el modelo de sustitución nucleotídica, pueden ser estimadas durante el proceso o bien pueden ser tomadas de un árbol NJ previamente construido. La elección del mejor árbol NJ de partida se realiza con el programa Moldtest3.7 (Posada & Crandall 1998), que ofrece la posibilidad de construir 56 árboles NJ basados en 56 modelos evolutivos diferentes y elegir entre ellos el que menos longitud total tiene. Los parámetros del modelo de sustitución nucleotídica que se estiman para el NJ elegido son los que se emplean para la construcción del árbol ML final.

Para construir el árbol filogenético basado en este método, se utilizó el programa PAUP4.0b10 (Swofford 2002). El árbol obtenido fue linealizado de acuerdo al modelo de Thorne (Thorne et al. 1998), que admite un ritmo de cambio variable a lo largo del tiempo y distinto para cada linaje, con el fin de conocer la antigüedad relativa de cada grupo.

### ***5.3 Construcción de un árbol filogenético basado en el método Bayesiano (BI, Bayesian Inference) (Huelsenbeck & Ronquist 2001)***

Este método supone un avance frente a otros métodos similares, como el *Maximum Likelihood* comentado anteriormente, ya que modifica los valores de las variables que definen el modelo de sustitución nucleotídica durante el proceso de construcción del árbol filogenético, de esta forma va creando árboles, que va mejorando en función de un valor de probabilidad. Para ello comienza construyendo un árbol a partir de una matriz de secuencias de ADN y teniendo en cuenta el modelo de

sustitución nucleotídica programado por el usuario. A continuación calcula la probabilidad de ese árbol y a partir de este punto se repite el proceso inicial un número determinado de veces (generaciones). En cada generación se modifican los parámetros iniciales del modelo de sustitución nucleotídica con el fin de mejorar el valor de probabilidad y por tanto el resultado final.

Gracias a este método obtenemos un árbol que va a tener en cada uno de sus nodos un valor de probabilidad o *bootstrap*, que nos va a indicar la confianza o veracidad de esas uniones. De esta forma se puede saber que nodos son más fiables y cuáles no. Cuanto más alto es el valor de *bootstrap*, mayor probabilidad hay de que esa unión sea un reflejo de la realidad. Mientras que valores bajos indicarían que esos nodos se habrían creado simplemente por azar.

Para construir el árbol filogenético basado en este método, se utilizó el programa MrBayes (Huelsenbeck & Ronquist 2001).



# RESULTADOS



### 1. CARACTERIZACIÓN DE LAS SECUENCIAS DE ADN DEL INTRÓN 2 DE LOS GENES MHC DE CLASE I

11 especies pertenecientes al género *Serinus* han sido analizadas en este proyecto de tesis (Tabla M1). A partir de muestras de sangre se ha extraído, amplificado y secuenciado el fragmento de ADN correspondiente al intrón 2 de los genes MHC de clase I. Para completar el estudio se han añadido al análisis las secuencias del intrón 2 del pinzón vulgar (*Fringilla coelebs*) y del pollo (*Gallus gallus*). Las secuencias obtenidas tenían una longitud media de 309 pb. La media de todo el grupo, incluidas las secuencias del pinzón vulgar y el pollo fue de 299 pb.

Todas las secuencias fueron alineadas con el programa MEGA 7 (Kumar et al. 2016), como se indica en el apartado 4.1 de Materiales y Métodos. De esta manera se puede observar la homología a nivel de nucleótidos. El alineamiento de dichas secuencias se puede ver con detalle en la Figura R1.

Lo más destacable de la Figura R1 es la cantidad de posiciones nucleotídicas conservadas entre las diferentes especies del género *Serinus*. Sin tener en cuenta al canario silvestre (*Serinus canaria*), las secuencias del intrón 2 del resto de especies del género *Serinus* sólo se diferencian en 28 posiciones de un total de 309 de media, mientras que el resto permanecen conservadas. El canario silvestre (*Serinus canaria*) es la excepción del grupo de los canarios, ya que posee 183 posiciones diferentes con respecto al resto de individuos de su mismo género, además de que su secuencia es algo más corta que la del resto, contando sólo con 286 nucleótidos (ver Figura R1).

La secuencia del intrón 2 del pinzón vulgar (*Fringilla coelebs*) cuenta con un total de 292 nucleótidos, de los cuales 198 son diferentes al resto del grupo de estudio. Por último la secuencia del intrón 2 del pollo (*Gallus gallus*) tiene un total de 229 nucleótidos, de los cuales 173 son diferentes al resto del grupo (ver figura R1).



| Especies                      | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
|-------------------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| <i>Serinus canaria</i>        | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | g  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | g  | a  | g  | a  | t  | g  | g  | g  |
| <i>Serinus sulphuratus</i>    | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus atrogularis</i>    | g | t | g | a | g | t | g | c | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus citrinelloides</i> | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus citrinipectus</i>  | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus dorsostratus</i>   | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus flaviventris</i>   | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus gularis</i>        | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus mozambicus</i>     | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus striolatus</i>     | g | t | g | a | g | t | g | g | g | g  | g  | g  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Serinus thibetanus</i>     | g | t | g | a | g | t | g | g | g | g  | g  | a  | a  | t  | c  | t  | g  | t  | g  | c  | a  | g  | t  | t  | g  | g  | a  | a  | t  | g  | g  | g  |
| <i>Gallus gallus</i>          | g | t | g | a | g | c | a | c | g | g  | c  | c  | g  | g  | g  | g  | c  | c  | g  | c  | g  | g  | c  | t  | c  | c  | g  | t  | g  | g  | g  | t  |
| <i>Fringilla coelebs</i>      | g | t | g | a | g | t | g | g | g | g  | g  | g  | g  | t  | t  | c  | t  | g  | t  | g  | c  | a  | g  | c  | t  | g  | g  | a  | a  | t  | g  | g  |

| 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| a  | t  | c  | t  | g  | t  | a  | g  | g  | t  | c  | t  | t  | g  | g  | a  | t  | g  | t  | g  | t  | a  | t  | t  | c  | a  | c  | a  | a  | g  | g  | c  | t  | c  | c  | a  | a  |
| a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | g  | g  | t  | g  | g  | a  | g  | a  | t  | c  | a  | g  | t  | g  | g  | g  | g  | c  | t  | g  | g  | g  | g  |
| a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | g  | t  | t  | g  | g  | a  | g  | a  | t  | c  | a  | g  | t  | g  | g  | g  | g  | c  | t  | g  | g  | g  | g  |
| a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | g  | g  | t  | g  | g  | a  | g  | a  | t  | c  | a  | g  | t  | g  | g  | g  | g  | c  | t  | g  | g  | g  | c  |
| a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | g  | g  | t  | g  | g  | a  | g  | a  | t  | c  | a  | g  | t  | g  | g  | g  | g  | c  | t  | g  | g  | g  | g  |
| a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | g  | g  | t  | g  | g  | a  | g  | a  | t  | c  | a  | g  | t  | g  | g  | g  | g  | c  | t  | g  | g  | g  | g  |
| a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | g  | g  | t  | g  | g  | a  | g  | a  | t  | c  | a  | g  | t  | g  | g  | g  | g  | c  | t  | g  | g  | g  | g  |
| a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | g  | g  | t  | g  | g  | a  | g  | a  | t  | c  | a  | g  | t  | g  | g  | g  | g  | c  | t  | g  | g  | g  | g  |
| g  | t  | g  | g  | g  | a  | t  | g  | g  | g  | c  | t  | c  | c  | a  | t  | g  | g  | c  | g  | c  | a  | g  | t  | g  | c  | c  | g  | c  | c  | c  | a  | c  | a  | c  | c  | c  |
| a  | a  | a  | c  | t  | g  | t  | g  | g  | g  | g  | c  | t  | t  | g  | t  | g  | t  | g  | a  | g  | a  | t  | t  | g  | a  | t  | g  | g  | g  | g  | g  | c  | t  | t  | g  | g  |

| 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|
| g  | g  | g  | a  | g  | t  | g  | a  | a  | a  | t  | g  | g  | g  | a  | a  | t  | c  | t  | a  | t  | c  | a  | g  | g  | c  | t  | g  | g  | a  | t   | g   | g   | g   | a   | t   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | t  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | t  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | g  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | t  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | c  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | c  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | c  | a  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | t  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | t  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | t  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | g  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | t  | a  | c   | a   | g   | g   | g   | c   |
| t  | g  | g  | g  | a  | t  | c  | t  | a  | t  | a  | g  | g  | t  | c  | t  | c  | a  | g  | a  | t  | g  | c  | a  | g  | a  | t  | t  | c  | a  | c   | a   | g   | g   | g   | c   |
| c  | c  | c  | a  | g  | g  | c  | c  | t  | g  | g  | c  | c  | c  | t  | g  | c  | c  | c  | g  | g  | c  | g  | g  | c  | a  | c  | c  | g  | t  | c   | c   | c   | g   | g   | g   |
| a  | t  | g  | g  | g  | a  | t  | c  | t  | c  | t  | a  | g  | g  | t  | c  | t  | c  | g  | a  | t  | a  | c  | g  | g  | a  | t  | t  | c  | a  | c   | a   | g   | g   | g   | g   |

| 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| g   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | t   | a   | t   | g   | g   | g   | t   | t   | g   | g   | g   | a   | t   | g   | a   | g   | a   | t   | c   | t   |
| t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | g   | a   | t   | c   | c   | a   | t   | t   | a   | g   | g   | c   |
| t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | g   | a   | t   | c   | c   | a   | t   | t   | a   | g   | g   | c   |
| t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | g   | a   | t   | c   | c   | a   | t   | t   | a   | g   | g   | c   |
| t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | t   | a   | t   | c   | t   | a   | t   | t   | a   | g   | g   | c   |
| t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | g   | a   | t   | c   | c   | a   | t   | t   | a   | g   | g   | c   |
| t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | t   | a   | t   | c   | t   | a   | t   | t   | a   | g   | g   | c   |
| t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | g   | a   | t   | c   | c   | a   | t   | t   | a   | g   | g   | c   |
| g   | c   | t   | g   | c   | c   | c   | g   | t   | c   | a   | c   | a   | g   | c   | c   | c   | c   | a   | c   | c   | g   | c   | g   | c   | t   | c   | g   | g   | g   | g   |
| c   | t   | c   | t   | g   | a   | g   | g   | g   | g   | c   | t   | g   | a   | a   | a   | t   | g   | g   | g   | g   | t   | t   | c   | t   | g   | t   | t   | a   | g   | g   |

| 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| g   | t   | g   | g   | g   | g   | t   | t   | g   | g   | a   | t   | t   | g   | g   | g   | a   | t   | c   | c   | a   | t   | g   | g   | g   | g   | c   | t   | c   | c   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | a   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | a   | g   | g   | t   | t   |
| t   | g   | c   | c   | g   | c   | g   | t   | c   | c   | c   | g   | g   | g   | g   | g   | g   | a   | c   | c   | c   | c   | a   | a   | c   | c   | c   | a   | t   | c   |
| c   | t   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   | t   | c   | t   | t   | g   | t   | g   | a   | t   | c   | c   | a   | t   | g   | t   | g   | g   |

| 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| a   | a   | g   | t   | g   | g   | c   | t   | g   | g   | a   | a   | t   | g   | a   | g   | g   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | c   | t   | c   |
| g   | g   | g   | a   | t   | g   | t   | g   | g   | a   | t   | t   | g   | a   | t   | g   | t   | g   | g   | c   | t   | c   | c   | g   | a   | g   | g   | g   | g   | c   |
| g   | g   | g   | a   | t   | g   | t   | g   | g   | a   | t   | t   | g   | a   | t   | g   | t   | g   | g   | c   | t   | c   | t   | g   | a   | g   | g   | g   | g   | c   |
| g   | g   | g   | a   | t   | g   | t   | g   | g   | a   | t   | t   | g   | a   | t   | g   | t   | g   | g   | c   | t   | c   | c   | g   | a   | g   | g   | g   | g   | c   |
| g   | g   | g   | a   | t   | g   | t   | g   | g   | a   | t   | t   | g   | a   | t   | g   | t   | g   | g   | c   | t   | c   | c   | g   | a   | g   | g   | g   | g   | c   |
| g   | g   | g   | a   | t   | g   | t   | g   | g   | a   | t   | t   | g   | a   | t   | g   | t   | g   | g   | c   | t   | c   | c   | g   | a   | g   | g   | g   | g   | c   |
| g   | g   | g   | a   | t   | g   | t   | g   | g   | a   | t   | t   | g   | a   | t   | g   | t   | g   | g   | c   | t   | c   | c   | g   | a   | g   | g   | g   | g   | c   |
| g   | g   | g   | a   | t   | g   | t   | g   | g   | a   | t   | t   | g   | a   | t   | g   | t   | g   | g   | c   | t   | c   | c   | a   | a   | g   | g   | g   | g   | c   |
| c   | c   | c   | g   | c   | t   | g   | c   | a   | g   | t   | g   | g   | g   | a   | g   | c   | c   | c   | c   | g   | g   | a   | g   | c   | c   | g   | g   | a   | g   |
| a   | t   | t   | g   | a   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | a   | g   | g   | g   | g   | c   | t   | g   | g   | g   | a   | t   | g   | a   | a   |

| 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| c   | c   | t   | c   | a   | g   | a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   |     | a   | c   | a   | g   | a   | g   | c   | t   | g   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | a   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   | g   | t   | c   | t   | c   | g   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | a   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | g   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | c   | c   | t   | c   | t   | g   | g   | g   | g   | a   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| t   | g   | g   | g   | a   | t   | g   | a   | a   | g   | c   | t   | c   | t   | g   | g   | g   | g   | a   |     | t   | c   | t   | c   | c   | a   | t   | c   | a   | g   |
| g   | g   | g   | c   | c   | c   | c   | t   | c   | a   | c   | c   | c   | c   | c   | t   | g   | c   | c   | c   | g   | g   | c   | t   | g   | t   | g   | t   | t   | t   |
| g   | c   | t   | c   | t   | g   | g   | g   | g   | g   | t   | c   | c   | c   | c   | a   | t   | c   | a   | g   | a   | c   | t   | c   | c   | a   | t   | c   | g   | g   |

| 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| a   | a   | t   | g   | g   | g   | g   | c   | t   | c   | t   | g   | t   | g   | g   | g   | g   | c   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | c   | a   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| a   | c   | t   | c   | t   | g   | g   | g   | g   | c   | t   | c   | c   | a   | t   | a   | g   | g   | g   | c   | t   | g   | g   | a   | a   | t   | t   | t   | g   | g   |
| c   | a   | g   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| g   | c   | t   | g   | g   | a   | a   | t   | g   | g   | g   | g   | c   | t   | t   | t   | g   | c   | a   | g   | g   | g   | a   | t   | g   | g   | g   | a   | t   | a   |

| 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| c   | c   | a   | t   | g   | g   | g   | t   | c   | t   | c   | t   | g   | t   | g   | g   | a   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | a   |
| c   | t   | t   | t   | g   | t   | g   | g   | g   | g   | a   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
| c   | t   | t   | t   | g   | t   | g   | t   | g   | g   | c   | t   | g   | g   | g   | a   | c   | a   | c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | a   | g   |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| c   | a   | a   | t   | t   | c   | c   | a   | t   | g   | g   | g   | t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | a   | a   | a   | t   | c   | c   |

| 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| g   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
| t   | c   | t   | c   | t   | g   | t   | g   | g   | g   | t   | a   | c   | a   | a   | t   | c   | c   | c   | c   | c   | c   | a   | g   |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| c   | c   | c   | c   | a   | g   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Figura R1:  
Alineamiento de las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio. En azul claro se pueden ver las posiciones conservadas entre las distintas especies mientras que en azul oscuro se muestran las posiciones diferentes.

## 1.1 Frecuencias nucleotídicas de las secuencias del intrón 2

Para conocer la composición nucleotídica de las secuencias de estudio se calcularon las frecuencias de los cuatro tipos de bases que componen el ADN. En la Tabla R1 se pueden ver con detalle las frecuencias nucleotídicas obtenidas en las 13 especies incluidas en este análisis.

Tabla R1: Frecuencias nucleotídicas de las secuencias del intrón 2 de los genes MHC de clase I en las especies de estudio.

| Especies                      | T (%) | C (%) | A (%) | G (%) | Nucleótidos totales |
|-------------------------------|-------|-------|-------|-------|---------------------|
| <i>Serinus canaria</i>        | 25,5  | 17,8  | 18,5  | 38,1  | 286                 |
| <i>Serinus sulphuratus</i>    | 25,2  | 16,8  | 18,8  | 39,2  | 309                 |
| <i>Serinus atroglaris</i>     | 25,9  | 16,8  | 18,8  | 38,5  | 309                 |
| <i>Serinus citrinelloides</i> | 25,2  | 17,2  | 18,4  | 39,2  | 309                 |
| <i>Serinus citrinipectus</i>  | 25,5  | 16,5  | 19,0  | 39,0  | 310                 |
| <i>Serinus dorsostriatus</i>  | 24,9  | 17,5  | 19,4  | 38,2  | 309                 |
| <i>Serinus flaviventris</i>   | 25,2  | 18,1  | 18,8  | 37,9  | 309                 |
| <i>Serinus gularis</i>        | 25,2  | 16,8  | 18,8  | 39,2  | 309                 |
| <i>Serinus mozambicus</i>     | 25,9  | 16,5  | 19,4  | 38,2  | 309                 |
| <i>Serinus striolatus</i>     | 24,9  | 17,2  | 18,8  | 39,2  | 309                 |
| <i>Serinus thibetanus</i>     | 25,6  | 17,2  | 19,4  | 37,9  | 309                 |
| <i>Gallus gallus</i>          | 11,8  | 42,4  | 9,6   | 36,2  | 229                 |
| <i>Fringilla coelebs</i>      | 25,7  | 17,1  | 18,2  | 39,0  | 292                 |
| <b>Frecuencia media</b>       | 24,6  | 18,6  | 18,3  | 38,5  | 299                 |

Como se puede ver en la Tabla R1, las aves que pertenecen al género *Serinus* muestran una composición nucleotídica con valores muy similares entre ellas. El pinzón vulgar (*Fringilla coelebs*) por su parte, también muestra valores parecidos, sin embargo el pollo (*Gallus gallus*) destaca por sus porcentajes de Timina, Citosina y Adenina. El porcentaje de Timina del pollo es de 11,8%, mientras que el del resto del grupo está alrededor del 25%. En el caso del porcentaje de Citosina, el pollo tiene un 42,4% de la misma, mientras que el resto de las aves sólo tienen un 18% aproximadamente. Por último destacar el bajo porcentaje de Adenina que posee el pollo (9,6%), con respecto al 38% del resto de especies.

Tabla R2: Frecuencias de Guanina + Citosina (G+C) en las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio.

| Especies                      | G (%) | C (%) | G+C (%) |
|-------------------------------|-------|-------|---------|
| <i>Serinus canaria</i>        | 38,1  | 17,8  | 55,9    |
| <i>Serinus sulphuratus</i>    | 39,2  | 16,8  | 56,0    |
| <i>Serinus atroglaris</i>     | 38,5  | 16,8  | 55,3    |
| <i>Serinus citrinelloides</i> | 39,2  | 17,2  | 56,3    |
| <i>Serinus citrinipectus</i>  | 39,0  | 16,5  | 55,5    |
| <i>Serinus dorsostriatus</i>  | 38,2  | 17,5  | 55,7    |
| <i>Serinus flaviventris</i>   | 37,9  | 18,1  | 56,0    |
| <i>Serinus gularis</i>        | 39,2  | 16,8  | 56,0    |
| <i>Serinus mozambicus</i>     | 38,2  | 16,5  | 54,7    |
| <i>Serinus striolatus</i>     | 39,2  | 17,2  | 56,3    |
| <i>Serinus thibetanus</i>     | 37,9  | 17,2  | 55,0    |
| <i>Gallus gallus</i>          | 36,2  | 42,4  | 78,6    |
| <i>Fringilla coelebs</i>      | 39,0  | 17,1  | 56,2    |
| <b>Frecuencia media</b>       | 38,5  | 18,6  | 57,1    |

Como se aprecia en la Tabla R2, todas las secuencias del intrón 2 de las especies pertenecientes al género *Serinus* incluidas en este estudio tienen un porcentaje de Guanina + Citosina (G+C) muy similar, siendo siempre superior al 55%, a excepción de *Serinus mozambicus* que se queda en un 54,7%. El valor más destacable de esta tabla es el del pollo, con un 78,6% de G+C, lo que supone un 20% más que el resto de especies. Este aumento se debe al alto porcentaje de citosina que muestra esta especie (ver Tablas R1 y R2).

## 1.2 Distancias genéticas entre las secuencias del intrón 2

Para analizar el grado de conservación de las secuencias del intrón 2 de los genes MHC de clase I en nuestras aves de estudio, se calcularon las distancias genéticas existentes entre dichas secuencias. De esta manera se puede ver el grado de divergencia o similitud existente entre nuestras especies de estudio. Esta distancia genética puede tomar valores que oscilan entre 0 y 1, siendo 0 el valor más pequeño y por tanto el dato

que nos indicaría que nuestro par de secuencias son iguales, y 1 el valor más alto, que indicaría la mayor divergencia observada entre dos secuencias. Como se ha comentado anteriormente, esta distancia relaciona el número de sitios variables con el número total de nucleótidos de dos secuencias. En la Tabla R3 se puede ver la matriz de distancias genéticas de las secuencias del intrón 2 de los genes MHC de clase I de las aves pertenecientes al género *Serinus* utilizadas en este estudio. Además se añadieron a la comparación las secuencias de *Fringilla coelebs* y *Gallus gallus*.



Tabla R3: Matriz de distancias genéticas entre las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio

|    | Especies                      | 1            | 2            | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           | 11           | 12           |
|----|-------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1  | <i>Serinus canaria</i>        |              |              |              |              |              |              |              |              |              |              |              |              |
| 2  | <i>Serinus sulphuratus</i>    | <b>0,633</b> |              |              |              |              |              |              |              |              |              |              |              |
| 3  | <i>Serinus atrogularis</i>    | <b>0,633</b> | 0,010        |              |              |              |              |              |              |              |              |              |              |
| 4  | <i>Serinus citrinelloides</i> | <b>0,636</b> | 0,006        | 0,016        |              |              |              |              |              |              |              |              |              |
| 5  | <i>Serinus citrinipectus</i>  | <b>0,629</b> | <b>0,220</b> | <b>0,230</b> | <b>0,227</b> |              |              |              |              |              |              |              |              |
| 6  | <i>Serinus dorsostriatus</i>  | <b>0,633</b> | 0,013        | 0,023        | 0,019        | <b>0,223</b> |              |              |              |              |              |              |              |
| 7  | <i>Serinus flaviventris</i>   | <b>0,640</b> | 0,016        | 0,026        | 0,023        | <b>0,236</b> | 0,029        |              |              |              |              |              |              |
| 8  | <i>Serinus gularis</i>        | <b>0,633</b> | <b>0,000</b> | 0,010        | 0,006        | <b>0,220</b> | 0,013        | 0,016        |              |              |              |              |              |
| 9  | <i>Serinus mozambicus</i>     | <b>0,626</b> | 0,013        | 0,023        | 0,019        | <b>0,207</b> | 0,026        | 0,029        | 0,013        |              |              |              |              |
| 10 | <i>Serinus striolatus</i>     | <b>0,629</b> | 0,013        | 0,023        | 0,013        | <b>0,233</b> | 0,026        | 0,029        | 0,013        | 0,026        |              |              |              |
| 11 | <i>Serinus thibetanus</i>     | <b>0,636</b> | 0,023        | 0,032        | 0,029        | <b>0,233</b> | 0,029        | 0,039        | 0,023        | 0,036        | 0,029        |              |              |
| 12 | <i>Gallus gallus</i>          | <b>0,734</b> | <b>0,742</b> | <b>0,734</b> | <b>0,738</b> | <b>0,751</b> | <b>0,742</b> | <b>0,742</b> | <b>0,742</b> | <b>0,742</b> | <b>0,747</b> | <b>0,751</b> |              |
| 13 | <i>Fringilla coelebs</i>      | <b>0,699</b> | <b>0,712</b> | <b>0,712</b> | <b>0,716</b> | <b>0,695</b> | <b>0,719</b> | <b>0,719</b> | <b>0,712</b> | <b>0,723</b> | <b>0,712</b> | <b>0,719</b> | <b>0,790</b> |

Sombreado en color amarillo se puede ver las distancias genéticas entre el canario silvestre (*Serinus canaria*) y el resto de especies del análisis. La media de estos valores es de 0,646. Este dato es destacable ya que se trata de especies pertenecen al mismo género y por lo tanto no alejadas genéticamente, al menos en teoría. Además, cuando esta especie es comparada con el pinzón vulgar (*Fringilla coelebs*) y el pollo (*Gallus gallus*), los valores de distancia genética aumentan hasta 0,699 y 0,734 (ver Tabla R3). Observando estos resultados, resulta llamativo que el canario silvestre este casi a la misma distancia de especies que pertenecen a su mismo género, que de dos especies más antiguas y aparentemente mucho más alejadas del mismo genéticamente.

Sombreado en rojo se pueden ver las distancias genéticas entre el pinzón vulgar (*Fringilla coelebs*) y el pollo (*Gallus gallus*) con respecto a las especies pertenecientes al género *Serinus*. Las distancias más grandes se observan entre el pollo y el resto de aves pertenecientes al género *Serinus*, valor a esperar si se tiene en cuenta que se trata de grupos diferentes, ya que pertenecen a diferentes órdenes, (orden *Galliformes* en el caso del pollo y orden *Passeriformes* en el caso del género *Serinus*).

En color morado se puede ver el único cero de toda la matriz. Este valor se da entre dos especies pertenecientes al género *Serinus*, el canario de cabeza estriada (*Serinus gularis*) y el canario de Brimstone (*Serinus sulphuratus*). Lo que nos muestra este dato es que las secuencias del intrón 2 de los genes MHC de clase I de estas dos especies son idénticas, tal y como se muestra en el alineamiento múltiple de la Figura R1. Sombreado en color azul, se puede ver la relación entre el canario de pecho limón (*Serinus citrinpectus*) y el resto de especies. Es de esperar que esta especie sea diferente al pinzón vulgar y al pollo, pero lo que más llama la atención es que comparado con el resto de especies del género *Serinus*, y a excepción del canario silvestre, la secuencia del intrón 2 del canario de pecho limón diverge en más posiciones nucleotídicas respecto al resto del grupo de los canarios.

Por último, destacar que el resto de especies del estudio que forman parte del género *Serinus*, (todas exceptuando *S. canaria* y *S. citrinpectus*) muestran valores muy bajos de distancia genética, lo que significa que sus secuencias son muy similares (ver Tabla R3).

Para ver de forma gráfica los resultados obtenidos en el cálculo de las distancias genéticas entre las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio, se recurrió a la realización de un árbol filogenético basado en dichas distancias.

En el caso de los genes del MHC y por lo tanto del intrón 2, no es útil realizar árboles filogenéticos para estudiar las relaciones evolutivas entre especies, debido a que tienen una forma muy peculiar de evolución. Sin embargo, en este caso es útil para obtener una representación gráfica de la similitud entre las secuencias del intrón 2, al margen de la relación filogenética existente entre las especies incluidas en el análisis.

En la Figura R2 se muestra el árbol filogenético *Neighbor-joining* basado en las distancias genéticas que se detallan en la Tabla R3. Este árbol muestra las mismas relaciones entre especies que se han establecido gracias al cálculo de las distancias genéticas. Así, se puede ver como la secuencia del pollo (*G. gallus*), es la que más diverge del resto del grupo, seguida por la del pinzón vulgar (*F. coelebs*). Dentro del grupo de los canarios, el que tendría la secuencia del intrón 2 más diferente sería el canario silvestre (*S. canaria*). El resto del grupo *Serinus*, formaría un grupo compacto en el que sólo destacaría la secuencia del canario de pecho limón, la cual quedaría fuera del mismo, aunque muy cerca.

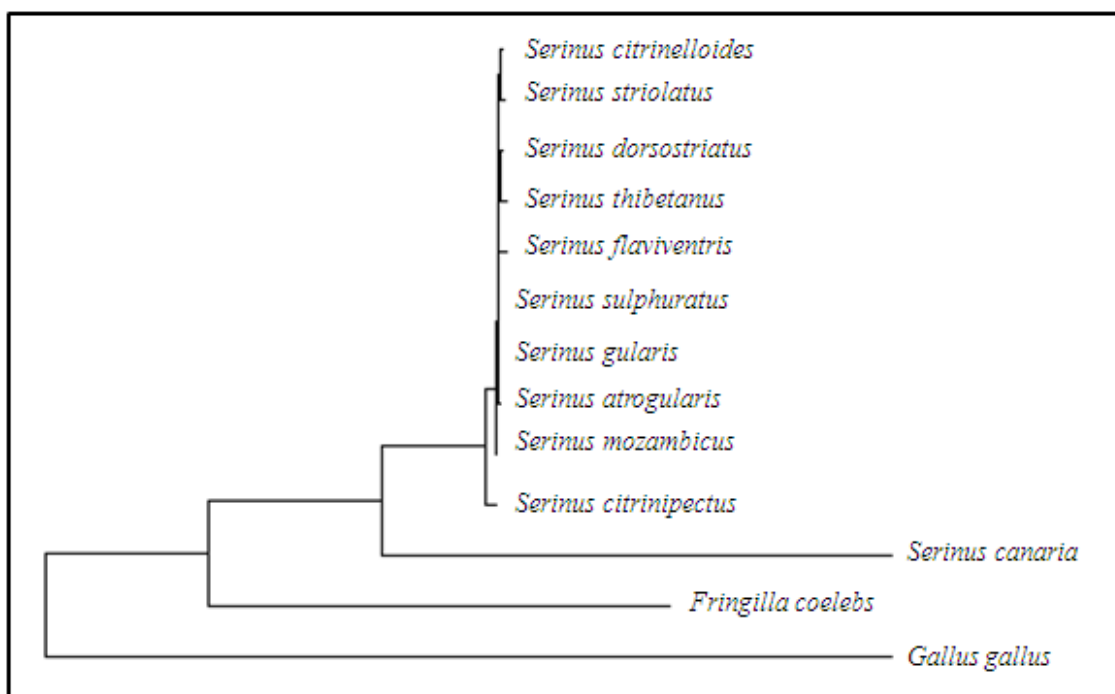


Figura R2: Árbol filogenético *Neighbor-joining* (NJ) basado en las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio.

### ***1.3 Porcentaje de similitud entre las secuencias del intrón 2***

Otra forma de obtener una medida aproximada del grado de conservación entre las secuencias del intrón 2 de los genes MHC de clase I es calculando el porcentaje de similitud existente entre ellas. Este análisis proporciona un apoyo al cálculo de las distancias genéticas, dando más fiabilidad al estudio de la conservación de la región del intrón 2. En la Tabla R4 se detallan los valores obtenidos del porcentaje de similitud entre las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio.

Sombreado en color rojo se muestran los porcentajes de similitud entre las secuencias del intrón 2 del pinzón vulgar (*F. coelebs*) y el pollo (*G. gallus*) con respecto a las secuencias del resto de especies. La secuencias del intrón 2 del pollo es la más diferente, mostrando un porcentaje de similitud siempre por debajo del 20%, excepto cuando se compara con la del canario silvestre (*S. canaria*), que llega hasta un 21,33% (ver Tabla R4). En el caso del pinzón vulgar (*F. coelebs*), este muestra un valor en torno al 27% de similitud con respecto al intrón 2 de las especies del género *Serinus*, exceptuando el 29,45% que se obtiene con el canario silvestre (*S. canaria*). El valor más bajo se puede observar entre el pinzón vulgar (*F. coelebs*) y el pollo (*G. gallus*), con un 16,44%. Estos resultados se correlacionan con los obtenidos en la Tabla R3 de distancias genéticas.

Sombreado en color amarillo se muestran los porcentajes de similitud de las secuencias del intrón 2 del canario silvestre con respecto a las secuencias del resto de especies de su mismo género. Es importante destacar los valores tan bajos de similitud (34,63% el más alto) teniendo en cuenta que se trata de especies del mismo género. Es de mencionar que los valores de similitud obtenidos entre esta especie y el resto de canarios de estudio son muy similares al valor obtenido cuando es comparada con el pinzón vulgar (ver Tabla R4). Esta relación se refleja también en la Tabla R3 de distancias genéticas.

Tabla R4: Porcentaje de similitud entre las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio

|    | Especies                      | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    |
|----|-------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1  | <i>Serinus canaria</i>        |       |       |       |       |       |       |       |       |       |       |       |       |
| 2  | <i>Serinus sulphuratus</i>    | 33,98 |       |       |       |       |       |       |       |       |       |       |       |
| 3  | <i>Serinus atrogularis</i>    | 33,98 | 99,03 |       |       |       |       |       |       |       |       |       |       |
| 4  | <i>Serinus citrinelloides</i> | 33,66 | 99,35 | 98,38 |       |       |       |       |       |       |       |       |       |
| 5  | <i>Serinus citrinipectus</i>  | 34,19 | 77,74 | 76,77 | 77,10 |       |       |       |       |       |       |       |       |
| 6  | <i>Serinus dorsostriatus</i>  | 33,98 | 98,71 | 97,73 | 98,06 | 77,42 |       |       |       |       |       |       |       |
| 7  | <i>Serinus flaviventris</i>   | 33,33 | 98,38 | 97,41 | 97,73 | 76,13 | 97,09 |       |       |       |       |       |       |
| 8  | <i>Serinus gularis</i>        | 33,98 | 100   | 99,03 | 99,35 | 77,74 | 98,71 | 98,38 |       |       |       |       |       |
| 9  | <i>Serinus mozambicus</i>     | 34,63 | 98,71 | 97,73 | 98,06 | 79,03 | 97,41 | 97,09 | 98,71 |       |       |       |       |
| 10 | <i>Serinus striolatus</i>     | 34,30 | 98,71 | 97,73 | 98,71 | 76,45 | 97,41 | 97,09 | 98,71 | 97,41 |       |       |       |
| 11 | <i>Serinus thibetanus</i>     | 33,66 | 97,73 | 96,76 | 97,09 | 76,45 | 97,09 | 96,12 | 97,73 | 96,44 | 97,09 |       |       |
| 12 | <i>Gallus gallus</i>          | 21,33 | 19,09 | 19,74 | 19,42 | 18,39 | 19,09 | 19,09 | 19,09 | 19,09 | 18,77 | 18,45 |       |
| 13 | <i>Fringilla coelebs</i>      | 29,45 | 27,18 | 27,18 | 26,86 | 28,71 | 26,54 | 26,54 | 27,18 | 26,21 | 27,18 | 26,54 | 16,44 |

En color morado se encuentra el único 100% del total de porcentajes de similitud. Este valor muestra que las secuencias del intrón 2 de los genes MHC de clase I del canario de cabeza estriada (*Serinus gularis*) y del canario de Brimstone (*Serinus sulphuratus*) son idénticas (Tabla R4), tal y como se reflejaba en la Tabla R3 de distancias genéticas.

Sombreado en color azul se puede ver la relación entre el canario de pecho limón (*Serinus citrinpectus*) y el resto de especies (Tabla R4). Al igual que pasaba con las distancias genéticas (ver Tabla R3), este canario muestra un valor diferente al resto de porcentajes que se pueden observar entre las especies del género *Serinus* (a excepción del canario silvestre, que se ha comentado con anterioridad). En este caso el porcentaje de similitud medio de esta especie con el resto está en torno al 77%.

Por último, comentar que el resto de especies del estudio que forman parte del género *Serinus*, a excepción del canario silvestre (*S. canaria*) y el canario de pecho limón (*S. citrinpectus*) muestran porcentajes de similitud elevados y similares entre especies (ver Tabla R4), lo que denota gran una alta similitud entre las secuencias del intrón 2 de los genes MHC de clase I de estas especies.

## **2. FILOGENIA MOLECULAR DEL GÉNERO *SERINUS* BASADA EN EL GEN DEL CITOCROMO B MITOCONDRIAL**

Para interpretar el proceso evolutivo al que han estado sometidas las especies de estudio y así conocer su relación filogenética, se recurrió al gen del citocromo b mitocondrial. En la Tabla M2 se detallan las muestras empleadas para este análisis, así como su procedencia y número de acceso de GenBank. En este caso no se usan los genes de Histocompatibilidad, debido a que evolucionan de forma peculiar y por lo tanto no están recomendados para analizar relaciones evolutivas entre especies. Las secuencias utilizadas forman parte del archivo del grupo que dirige el Dr. Arnaiz Villena (ver Tabla M2). Analizando la filogenia molecular basada en el gen mitocondrial de las especies de aves silvestres pertenecientes al género *Serinus*, se puede contrastar la relación evolutiva existente entre dichas especies con los resultados observados en los análisis de las secuencias del intrón 2 de los genes MHC de clase I y de esta forma ver si existe alguna correlación.

Para analizar la filogenia molecular del género *Serinus* se realizaron tres tipos de análisis diferentes con el fin de verificar los resultados obtenidos y así obtener unas conclusiones más fiables. Se construyeron tres árboles filogenéticos basados en tres métodos distintos: método *Neighbor-joining* (NJ) (Saitou & Nei 1987), método *Maximum Likelihood* (ML) (Felsenstein 1981) y método Bayesiano (BI, *Bayesian Inference*) (Huelsenbeck & Ronquist 2001).

## 2.1 Árbol filogenético basado en el método *Neighbor-joining* (NJ) (Saitou & Nei 1987)

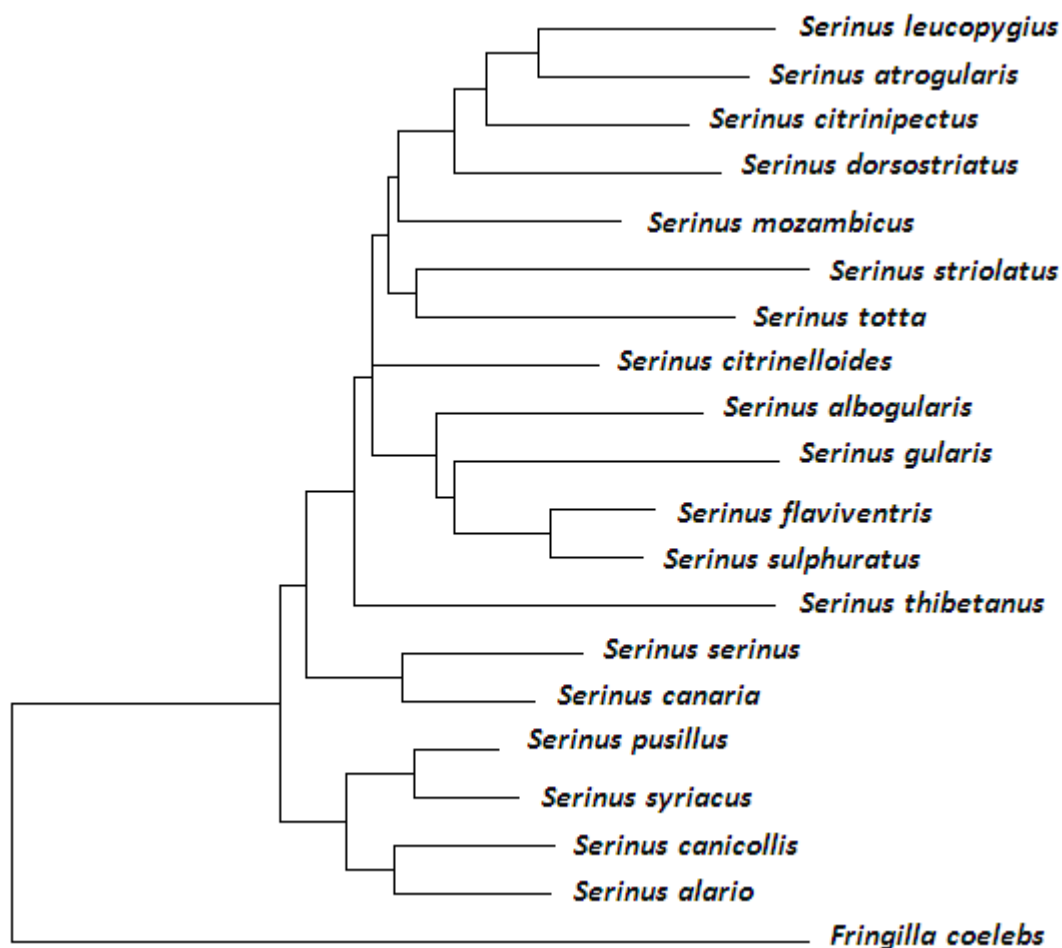


Figura R3: Árbol filogenético de las secuencias del gen del citocromo b mitocondrial de las aves de estudio obtenido mediante el método *Neighbor-joining*.

En la Figura R3 se puede observar el árbol filogenético obtenido mediante el método *Neighbor-joining* (Saitou & Nei 1987). Como se ha comentado anteriormente este método parte de una matriz de secuencias de ADN y analiza todos los árboles

resultantes de agrupar dos secuencias, dejando el resto desagrupadas. La longitud total de cada árbol es la suma de la longitud de cada una de sus ramas, siendo estas las distancias genéticas que separan los extremos de cada rama. El árbol resultante agrupa las secuencias de manera que se minimiza la distancia genética existente entre ellas. Este método es el menos complejo de los tres utilizados para analizar la filogenia de las aves del género *Serinus* y por lo tanto es el que nos da resultados menos fiables, ya que la única variable que tiene en cuenta es la distancia genética.

## 2.2 Árbol filogenético basado en el método Maximum Likelihood (ML) (Felsenstein 1981)

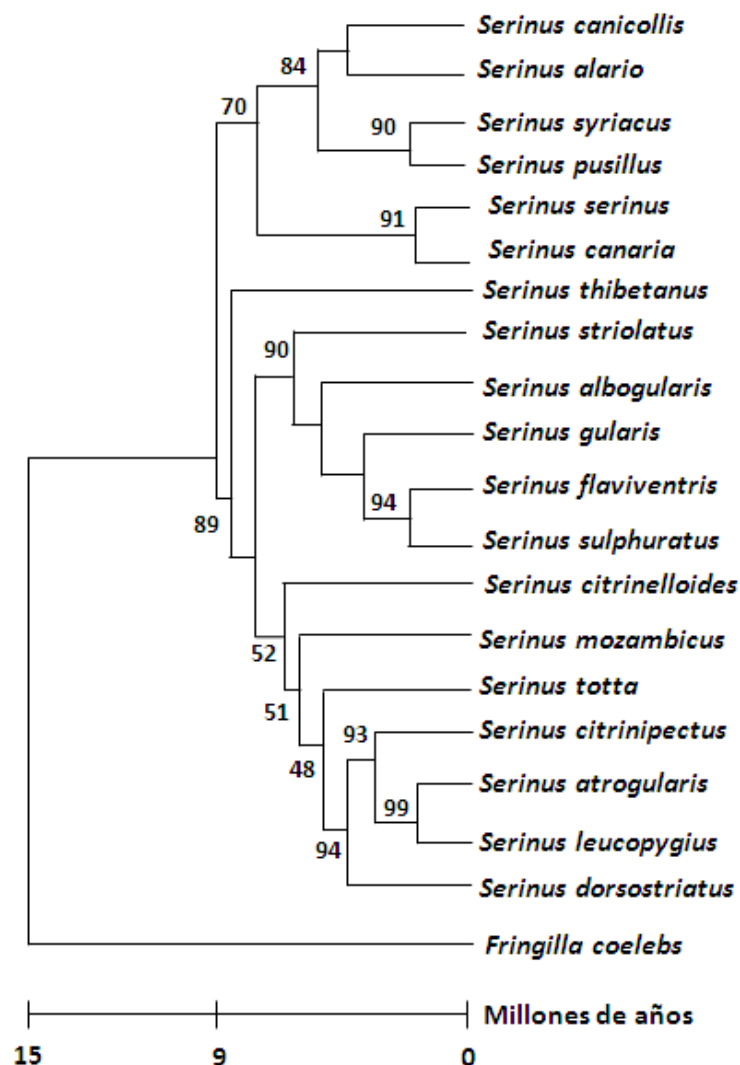


Figura R4: Árbol filogenético de las secuencias del gen del citocromo b mitocondrial de las aves de estudio obtenido mediante el método *Maximum Likelihood*.



En la Figura R4 se puede observar el árbol filogenético obtenido mediante el método *Maximum Likelihood* (ML) (Felsenstein 1981). Los valores numéricos que se sitúan en cada uno de los nodos indican el valor de “*bootstrap*” o fiabilidad de esas uniones. Así cuanto más alto sea este valor, hasta llegar a 100, más fiable será la unión de las especies de ese nodo. En el árbol representado en la Figura R4 se puede observar como todos los valores de “*bootstrap*” superan el número 70, a excepción de tres nodos que tendrían valores bajos de fiabilidad. El significado de los valores de “*bootstrap*” de todo el árbol se discutirá en profundidad en el apartado Discusión de este proyecto de tesis.

Como se ha comentado anteriormente este método parte de una matriz de secuencias de ADN a partir de la cual evalúa la probabilidad de obtener un árbol basándose en los valores que toman las variables que definen el modelo de sustitución nucleotídica. De tal forma que el árbol resultante será el de mayor probabilidad. El método *Maximum Likelihood* es más complejo y más fiable que el método *Neighbor-joining* ya que tiene en cuenta la manera de evolucionar que tienen las secuencias de ADN, mediante el uso de un modelo de sustitución nucleotídica.

Además mediante este método podemos linealizar el árbol de acuerdo al modelo de Thorne (Thorne et al. 1998), que admite un ritmo de cambio variable a lo largo del tiempo y distinto para cada linaje, con el fin de conocer la antigüedad relativa de cada grupo y establecer un tiempo aproximado de aparición de cada una de las especies del grupo de estudio.

### ***2.3 Árbol filogenético basado en el método Bayesiano (BI, Bayesian Inference) (Huelsenbeck & Ronquist 2001)***

En la Figura R5 se puede ver el árbol filogenético obtenido mediante el método Bayesiano (BI, Bayesian Inference) (Huelsenbeck & Ronquist 2001). Este método es aún más complejo y fiable que el *Maximum Likelihood* comentado anteriormente, ya que modifica los valores de las variables que definen el modelo de sustitución nucleotídica durante el proceso de construcción del propio árbol. El método comienza construyendo un árbol a partir de una matriz de secuencias de ADN y el modelo de sustitución nucleotídica programado por el usuario. A continuación calcula la probabilidad de ese árbol y a partir de este punto se repite el proceso inicial un número

determinado de veces. Para obtener el árbol de la Figura R5 se repitió el proceso durante 100000 generaciones, modificándose en cada una de ellas los parámetros iniciales del modelo de sustitución nucleotídica con el fin de mejorar el valor de probabilidad y por tanto el resultado final. Al igual que con el método *Maximum likelihood*, con el método Bayesiano se obtienen unos valores de “*bootstrap*” o fiabilidad en cada uno de los nodos del árbol, los cuales nos indican lo ciertas que son las relaciones entre las diferentes especies del grupo de estudio. En la Figura R5 se puede ver que los valores de “*bootstrap*” superan el número 75 excepto en tres nodos donde se pueden ver valores de 34, 40 y 47. Estos valores tan bajos de fiabilidad nos indican que estas uniones seguramente no sean del todo ciertas. En el apartado de Discusión, que se desarrollará a continuación, se comentará en profundidad el significado de los valores de “*bootstrap*” de la Figura R5.

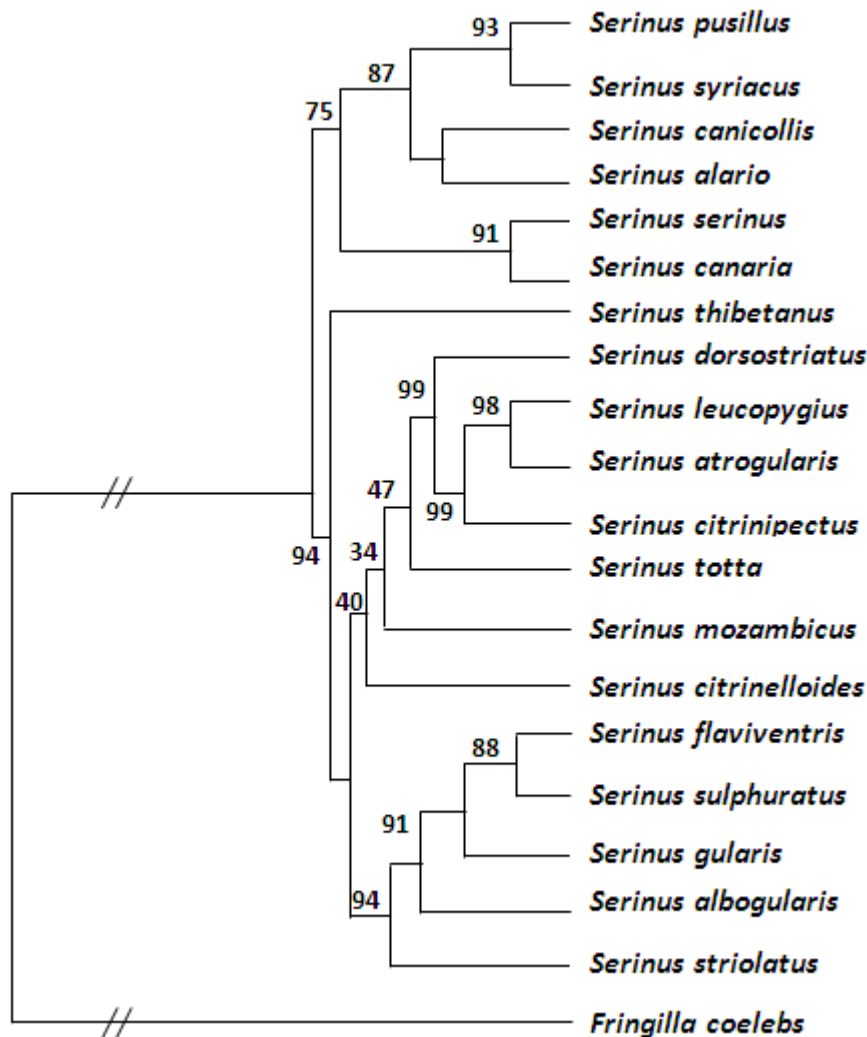


Figura R5: Árbol filogenético de las secuencias del gen del citocromo b mitocondrial de las aves de estudio obtenido mediante el método Bayesiano.



# DISCUSIÓN



## 1. FILOGENIA MOLECULAR Y EVOLUCIÓN DEL GÉNERO *SERINUS*

El género *Serinus* está formado por 45 especies (Sibley & Ahlquist 1990), las cuales son de pequeño tamaño y cuya distribución se centra principalmente África, aunque también se localizan en Europa occidental, Asia, Oriente Medio y los archipiélagos de la Macaronesia. Suelen ser de color verde, limón, marrón o gris estriado (Clement et al. 1993), como se puede ver en la descripción e ilustración de las especies utilizadas en este proyecto de tesis del apartado Introducción.

Para analizar la evolución filogenética del grupo de los canarios (género *Serinus*) se construyeron tres árboles filogenéticos mediante tres métodos diferentes (*Neighbor-joining*, *Maximum likelihood* y Bayesiano) basados en las secuencias del gen del citocromo b mitocondrial (ver Tabla M2) y así contrastar los resultados y comparar la evolución del gen mitocondrial con la evolución del Sistema Principal de Histocompatibilidad.

Los resultados obtenidos mediante los métodos de *Maximum likelihood* (Figura R4) y Bayesiano (Figura R5) son muy similares, lo que da peso a las relaciones mostradas por dichos árboles. Ambos árboles muestran nodos con valores altos de confianza estadística (valores de “bootstrap”) que agrupan a las mismas especies, mostrando incluso los valores más bajos de confianza (menores de 52) en los mismos nodos. Por otro lado, el árbol construido mediante el método *Neighbor-joining* (Figura R3) muestra casi las mismas relaciones entre especies que los otros dos métodos, a excepción de alguna que varía. Este método es menos complejo que los anteriores y no tiene en cuenta el modelo de sustitución nucleotídica, ni da valores de confianza o “bootstrap”, por lo que sólo es útil para obtener una visión rápida y aproximada de las relaciones entre las especies del género *Serinus*. Para conseguir resultados más fiables y precisos se suelen tomar como referencia los árboles obtenidos mediante los métodos *Maximum likelihood* y Bayesiano.

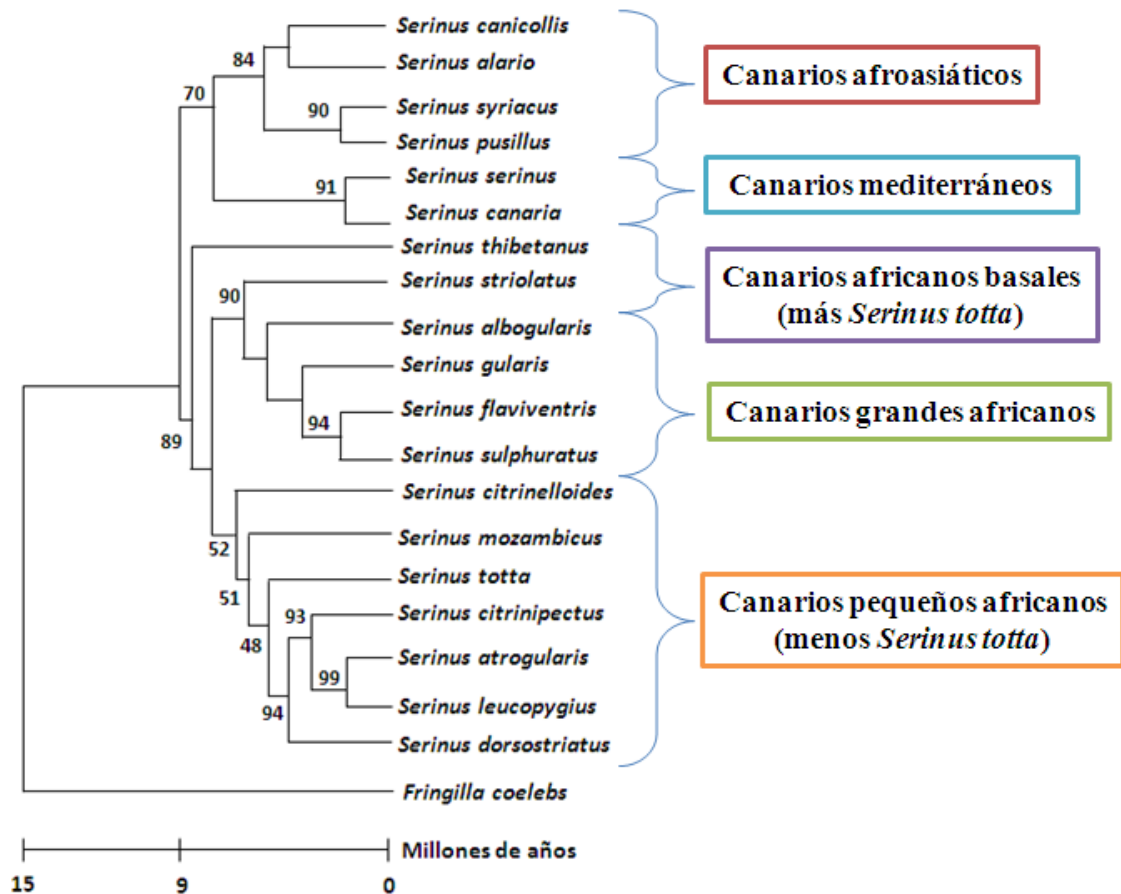


Figura D1: Árbol filogenético de las secuencias del gen del citocromo b mitocondrial de las aves de estudio obtenido mediante el método *Maximum likelihood* donde se especifican la clasificación de los canarios según su hábitat principal.

En la Figura D1 se puede observar un árbol filogenético construido mediante el método *Maximum likelihood* en el que se han señalado los grupos en los que se clasifican las aves pertenecientes al género *Serinus*. Si observamos esta figura, podemos ver que los canarios afroasiáticos aparecen agrupados con los canarios mediterráneos, con un valor de confianza de 70. Este valor es de 75 en el caso del árbol basado en el método Bayesiano de la Figura R5, lo que da fiabilidad a la unión de estos dos grupos de canarios. El grupo de los canarios afroasiáticos, formado por *Serinus canicollis*, *Serinus alario*, *Serinus syriacus* y *Serinus pusillus*, podría compartir ancestros comunes con el grupo de los canarios mediterráneos, el cual incluye a *Serinus serinus* y *Serinus canaria*, que podrían haber existido hace unos 7 millones de años, según los resultados obtenidos (ver Figuras R4, R5 y D1). La unión del canario silvestre (*S. canaria*) con el verdicillo (*S. serinus*), que ya fue propuesta con anterioridad (Clement et al. 1993), se ha confirmado en este trabajo con datos moleculares.

Otro grupo de canarios que se puede observar en la Figura D1 es el de los canarios grandes africanos, compuesto por *Serinus albogularis*, *Serinus gularis*, *Serinus flaviventris* y *Serinus sulphuratus*. Este grupo de canarios tiene un tamaño que oscila entre los 15-16 centímetros, característica que los diferencia del grupo de canarios pequeños africanos formado por *Serinus mozambicus*, *Serinus citrinipectus*, *Serinus citrinelloides*, *Serinus atrogularis*, *Serinus leucopygius* y *Serinus dorsostriatus*, cuyo tamaño está entre los 10 y 13 centímetros. Además estos dos grupos son diferentes en las características de su canto, siendo los canarios pequeños africanos los que tienen un canto variado y potente, mientras que los canarios grandes africanos poseen un canto menos potente y más monótono. Los valores de confianza obtenidos en los árboles de las Figuras R4 y R5 son similares, lo que da solidez a los resultados obtenidos.

Por último, en la Figura D1 se pueden ver tres especies que según estudios previos pertenecerían al grupo de los canarios africanos basales (Arnaiz-Villena et al. 2012). Dichas especies son, *Serinus totta*, *Serinus thibetanus* y *Serinus striolatus*. Sin embargo, en los análisis realizados en este proyecto de tesis, la especie de *Serinus totta* está incluida en el grupo de canarios pequeños africanos (ver Figuras R4, R5 y D1), a excepción del árbol construido con el método *Neighbor-joining* que aparece con *Serinus striolatus* (ver Figura R3). Esto quizás se deba a que faltan algunas especies por analizar del género *Serinus*. Es importante destacar que los valores de confianza del nodo que une a *Serinus totta* con el resto de los canarios pequeños africanos son menores de 50 tanto en el árbol basado en el método *Maximum likelihood* (Figura R4) como en el basado en el método Bayesiano (Figura R5), por lo que esta unión parece poco probable y por lo tanto no se debería de tener en cuenta. Por otro lado, los nodos que muestran a *Serinus striolatus* y *Serinus thibetanus* como canarios más basales o antiguos, poseen valores de confianza o “bootstrap” más elevados, por lo que pueden tomarse como probables dichas uniones (Figura D1).

En lo que respecta a la evolución del género *Serinus*, en la Figura D1 se puede ver como la edad aproximada de aparición de las aves de este género en la tierra es de unos 9 millones de años. Posteriormente, entre 1 y 2 millones de años después se dividieron en dos grupos principalmente. Uno de ellos dio lugar a los canarios africanos, que a su vez se dividió en tres grupos, los canarios africanos basales (*S. striolatus*, *S. totta* y *S. thibetanus*), que serían los más antiguos; los canarios grandes africanos (*S. sulphuratus*, *S. gularis*, *S. flaviventris* y *S. albogularis*); los canarios pequeños africanos



(*S. leucopygius*, *S. citrinelloides*, *S. dorsostratus*, *S. atrogularis*, *S. citrinipectus* y *S. mozambicus*). El otro grupo posiblemente dio lugar a los canarios que abandonaron África para colonizar otras regiones de la tierra. En este grupo se incluyen los canarios afroasiáticos (*S. canicollis*, *S. alario*, *S. syriacus* y *S. pusillus*) y los canarios mediterráneos (*S. serinus* y *S. canaria*). Estos resultados confirman estudios realizados con anterioridad (Arnaiz-Villena et al. 2012), a excepción del grupo de los canarios africanos basales, los cuales no quedan agrupados con valores altos de confianza en las Figuras R4, R5 y D1.

En la década de los 90 se postuló que los ancestros del género *Serinus* podían provenir de Eurasia (probablemente de Oriente Medio) (Clement et al. 1993), pero los resultados obtenidos en este proyecto y en trabajos anteriores (Zamora et al. 2006), apoyan la teoría de un origen africano, ya que el grupo más antiguo sería el de los canarios africanos (ver Figuras R4 y R5).

Gracias a los resultados obtenidos en el análisis de la filogenia molecular del género *Serinus* se puede describir la evolución del mismo a lo largo de millones de años. Si se comparan las Figuras R4 y R5 (basados en secuencias de citocromo b mitocondrial) con la Figura R2 (basado en secuencias del intrón 2 de los genes MHC de clase I), se puede apreciar como los genes de histocompatibilidad no son buenos para establecer relaciones filogenéticas entre especies, debido a su modo peculiar de evolución. En la Figura R2 se aprecia un grupo principal en el que están todas las especies del género *Serinus* a excepción del canario de pecho limón (*Serinus citrinipectus*) y del canario silvestre (*Serinus canaria*) que aparecen fuera de este grupo. El canario silvestre, sin embargo, formaría parte del grupo de canarios mediterráneos junto con el verdecillo (*Serinus serinus*) según los resultados obtenidos en los árboles basados en el citocromo b mitocondrial (Figuras R4 y R5). El canario de pecho limón (*Serinus citrinipectus*) por su parte, formaría parte del grupo de los canarios pequeños africanos (Figuras R4 y R5), relación que no se ve reflejada en el árbol de la Figura R2.

Por lo que las relaciones establecidas entre las secuencias del intrón 2 de los genes MHC de clase I reflejadas en la Figura R2, no siguen el mismo patrón de evolución que las secuencias del gen del citocromo b mitocondrial establecidas en las Figuras R4 y R5. Lo que puede indicar que los genes de histocompatibilidad siguen su propio patrón de evolución que puede estar influenciado por agentes externos tales como el ambiente y los patógenos existentes en su hábitat.

## **2. EL INTRÓN 2 DE LOS GENES MHC DE CLASE I EN LAS AVES SILVESTRES DEL GÉNERO *SERINUS***

Como se ha comentado en apartados anteriores, un intrón es un segmento de ADN que se localiza dentro de un gen codificante. El intrón no codifica proteínas ya que es eliminado del ARN mediante el proceso de corte y empalme conocido como “*splicing*”. Desde su descubrimiento en 1977, se ha investigado mucho sobre estos fragmentos de los cuales se pensaba inicialmente que eran ADN “basura”. Gracias a varios estudios se sabe que se localizan principalmente en organismos eucariotas (Gilson et al. 2006), donde su proporción varía desde unos pocos en organismos protistas (Mair et al. 2000) a una media de 8 intrones por gen en el caso de los seres humanos (Sakharkar et al. 2004).

En la actualidad, se está aceptando que los intrones poseen funciones importantes en los genomas y que no se trata solamente ADN “basura”. Entre estas funciones está su participación en la regulación de la transcripción y la organización del genoma (Chorev & Carmel 2012). Además se ha visto que la presencia de estos fragmentos de ADN ejerce una presión sustancial sobre el huésped que los contiene, ya que para su eliminación mediante el proceso de corte y empalme, se necesita un sistema complejo llamado espliceosoma, formado por más de 150 proteínas (Wahl et al. 2009). Este sistema tiene que ser codificado para que el proceso pueda llevarse a cabo, con el consiguiente gasto de tiempo y energía (Singh & Padgett 2009). Por otro lado, el reconocimiento de las secuencias de unión por parte del espliceosoma es vulnerable a mutaciones que pueden dañar el proceso de corte y empalme. Respecto a esto, se ha estimado que más del 50% de los trastornos genéticos en seres humanos están relacionados con la interrupción de las secuencias de reconocimiento por parte del espliceosoma (Lopez-Bigas et al. 2005).

Gracias a estas características, es de gran importancia el estudio de estos fragmentos de ADN tan peculiares. Por ello son el objeto principal de este proyecto de tesis.

### **2.1 Análisis de las secuencias del intrón 2 obtenidas**

Para el estudio de las secuencias del intrón 2 de los genes MHC de clase I se han analizado 11 especies diferentes pertenecientes al género *Serinus* (ver Tabla M1). A partir de muestras de sangre se ha obtenido el fragmento de ADN correspondiente al intrón 2 de los genes MHC de clase I. Además para completar el análisis de estas secuencias se han añadido las secuencias del pinzón vulgar (*Fringilla coelebs*) y del pollo (*Gallus gallus*). En la Figura R1 se puede ver el alineamiento realizado con las secuencias del intrón 2 de las especies del género *Serinus*, más el pinzón vulgar y el pollo.

Las secuencias del intrón 2 obtenidas de las especies del género *Serinus* tenían una longitud media de 309 pb, mientras que la secuencia del pinzón vulgar tiene 292 pb y el pollo 229 pb. El canario silvestre (*Serinus canaria*) destaca por tener un intrón de menor longitud respecto al resto del grupo de los canarios, ya que tiene sólo 286 pb. Si se analiza con detalle la Figura R1 se ve como el grupo de los *Serinus*, a excepción del canario silvestre, tiene un total de 281 posiciones conservadas en la secuencia del intrón 2, que como se ha mencionado anteriormente de media tiene una longitud de 309 pb. Por lo que esta secuencia está altamente conservada en los individuos analizados del grupo *Serinus*. Este hecho puede indicar que existe una presión positiva sobre esta secuencia que hace que no varíe a lo largo de millones de años. El caso del canario silvestre (*Serinus canaria*) es de destacar, ya que aun perteneciendo al grupo de los canarios, su secuencia tiene 183 posiciones diferentes al resto de un total de 286 pb. Este hecho podría deberse a las diferencias que existen entre los patógenos peninsulares y los de las islas donde habita este canario, además del propio efecto insular de aislamiento del resto de hábitats y ambientes que podría estar influyendo en la evolución del intrón 2 de este canario.

En lo que respecta al pinzón vulgar y al pollo, en la Figura R1 se puede ver cómo estas dos secuencias son muy diferentes a las secuencias del grupo de los canarios, siendo la del pollo la más dispar, ya que varía en 173 nucleótidos de un total de 229. Se trata de la secuencia más corta de todas las analizadas en este proyecto.

En la década de los 90, Kaufman y colaboradores denominaron al Sistema Principal de Histocompatibilidad del pollo “MHC mínimo esencial” (Kaufman et al. 1995; Kaufman et al. 1999), debido a las diferencias que estos observaron en el MHC del pollo al compararlo con otros MHC de mamíferos y el HLA humano. Estas diferencias serían principalmente, la proximidad de sus genes, la falta de genes de la

región MHC y el hecho de que posee intrones cortos. De la misma forma extendieron esta definición de “MHC mínimo esencial” al resto del grupo de las aves.

Como se puede ver en los resultados de este proyecto de tesis (Figura R1), la longitud de las secuencias de intrón 2 de los genes MHC de clase I del grupo de aves del género *Serinus* es mayor que la del pollo, de tal forma que no se podría hablar de “MHC mínimo esencial” en cuanto a la longitud de las secuencias en este grupo de aves *Passeriformes*.

Los organismos eucariotas tienen intrones con un tamaño, número y contexto de secuencia variable. El número y el tamaño reflejan la naturaleza y la eficiencia de la maquinaria del proceso de corte y empalme, el cual varía entre especies y entre diferentes linajes. Los parámetros del contexto son más complicados de analizar, ya que concierne parámetros como el contenido de la secuencia, la composición nucleotídica (por ejemplo el porcentaje de G+C), elementos transponibles y elementos funcionales (como potenciadores del proceso de corte y empalme).

Estudios de Wu y colaboradores (2013) muestran que el número y el tamaño de los intrones varían entre diferentes linajes. Dentro del grupo de los protozoos, donde se incluyen especies tales como, *Toxoplasma gondii* y *Plasmodium chabandi*, se pueden ver grandes diferencias, observándose intrones con un tamaño que varía desde  $10^1$  pb a  $3,16 \times 10^5$  pb (Wu et al. 2013b).

También observaron que en el grupo de los hongos y los invertebrados, la distribución del tamaño de los intrones es menos variada que en el caso de los protozoos. En ambos grupos, el tamaño medio de los intrones oscila entre  $10^1$  y  $10^3$  pb. *Ashbya gossypii loci* y *Saccharomyces cerevisiae* serían dos ejemplos de especies de hongos, mientras que *Drosophila melanogaster* y *Caenorhabditis elegans* serían dos ejemplos del grupo de los invertebrados (Wu et al. 2013b).

Por último vieron que los dos linajes donde se observa un tamaño del intrón más conservado son a su vez dos de los grupos más evolucionados. Se trata de las plantas y los vertebrados. El grupo de las plantas incluiría especies tales como, *Glycine max* y *Arabidopsis thaliana* y el grupo de los vertebrados incluiría especies como *Danio rerio*, *Xenopus tropicalis*, *Homo sapiens* y *Mus musculus*. Estos grupos tienen intrones cuyos tamaños oscilan entre  $10^2$  y  $10^3$  pb, y muestran un patrón de distribución muy similar en cuanto al tamaño y densidad de los intrones (Wu et al. 2013b).

Otros análisis realizados por Castillo Davis y colaboradores (2002), muestran como el tamaño de los intrones varía en función del nivel de expresión de los genes que

los contienen. Se ha observado que los intrones de genes que tienen un nivel alto de expresión son sustancialmente más cortos que aquellos que forman parte de genes con baja expresión. En humanos esta diferencia es bastante notable, ya que los intrones son de media 14 veces más cortos en los genes con alta expresión comparados con los de baja expresión. En organismos más simples como es el caso del nematodo *Caenorhabditis elegans*, esta diferencia es solamente el doble. En contraste, estos investigadores vieron que la densidad de los intrones en un gen no depende del nivel de expresión del mismo (Castillo-Davis et al. 2002). Una de las explicaciones que dan estos autores para entender el porqué de esta diferencia de tamaño, es la relacionada con el coste energético. Algunos de los intrones más largos se encuentran en los seres humanos, donde el orden de magnitud puede llegar a  $10^3$  pb. En los organismos eucariotas, se transcriben aproximadamente 20 nucleótidos por segundo, y el coste suele ser de 2 moléculas de ATP por nucleótido. Debido a esto, la selección natural estaría favoreciendo la selección de intrones cortos frente a los intrones más largos para los genes de alta expresión para así disminuir el coste energético (Castillo-Davis et al. 2002).

Otros estudios, como los realizados por Vinogradov (1999), se centran en analizar la relación entre el tamaño de los intrones con respecto al tamaño del genoma. Este autor analizó esta relación en un amplio rango evolutivo, que iba desde las levaduras hasta los seres humanos. Lo que pudo observar fue que a escala global, el tamaño promedio del intrón aumenta a medida que aumenta el tamaño del genoma. También observó algunas excepciones, como la de las levaduras de la especie *Saccharomyces cerevisiae*, las cuales tienen unos intrones con un tamaño extraordinariamente grande para el tamaño de su genoma. Esta característica también la observó en los tetrápodos (Vinogradov, 1999).

Hughes y Hughes (1995), estudiaron el tamaño de los intrones en el pollo y en el ser humano y vieron que los intrones del pollo eran significativamente más pequeños que sus homólogos en el ser humano. Por esta razón hipotetizaron que el decrecimiento en el tamaño del intrón y en el tamaño del genoma en aves con respecto a los mamíferos era debido a la demanda metabólica exigida por el vuelo (Hughes y Hughes 1995). Sin embargo los análisis realizados por Vinogradov (1999), no rebelaron diferencias significativas entre el tamaño de los intrones del pollo y los mamíferos, incluso usando los mismos datos de partida que Hughes y Hughes (1995). Por lo que no se puede establecer una conclusión firme sobre si los intrones de las aves son significativamente

más cortos que sus homólogos en mamíferos y por lo tanto no se puede establecer una conclusión respecto a la medida en que estos intrones están sujetos a presiones de selección impuestas por el vuelo.

Para intentar resolver estas dudas sobre la incidencia de procesos de alto requerimiento metabólico, como puede ser el vuelo, sobre el tamaño de los intrones en las aves, Waltari y Edwards (2002), analizaron el tamaño de 14 intrones en una especie pariente de las aves actuales, el cocodrilo americano (*Alligator mississippiensis*) y en 19 aves voladoras y no voladoras pertenecientes a 12 órdenes taxonómicos diferentes. Sus resultados indican que una fracción sustancial (alrededor del 66%) de la reducción del tamaño de los intrones y del tamaño del genoma ya se había dado en arcosaurios no voladores, por lo que con los análisis realizados por estos autores no se puede afirmar que el vuelo sea la causa de la reducción del tamaño de los intrones y del genoma. Lo que si confirmaron con sus resultados fue el hecho de que los genomas de las aves son más pequeños que los de los reptiles y mamíferos (Waltari & Edwards 2002).

Los investigadores Zhang y Edwards (2012), tras los estudios publicados de Hughes y Hughes (1995), Vinogradov (1999) y Waltari y Edwards (2002), decidieron continuar con el análisis del tamaño de los intrones y su posible relación con el vuelo. Identificaron un total de 12506 intrones en especies que incluían: el ser humano (*Homo sapiens*), ratón (*Mus musculus*), pequeño murciélago (*Myotis lucifugus*), pollo (*Gallus gallus*), marsupial colicorto gris (*Monodelphis domestica*), ornitorrinco (*Ornithorhynchus anatinus*), pavo (*Meleagris gallopavo*), y al diamante mandarín (*Taeniopygia guttata*) entre otras. Sus resultados muestran que los intrones de las aves y los reptiles son más cortos que los de los mamíferos y que esta diferencia es comparable con la exhibida por el tamaño del genoma. También vieron que dentro de los mamíferos, las especies voladoras como los murciélagos tenían alrededor del 70% de sus intrones más cortos que sus parientes más cercanos no voladores. En el grupo de las aves analizadas, vieron que las especies voladoras tenían un 60% de sus intrones más cortos que las especies no voladoras. Su estudio demuestra que existe un patrón complejo de evolución del tamaño del intrón y sugiere que las fuerzas de selección natural y las mutaciones varían entre los intrones de un mismo gen y entre especies. Estos autores demostraron la influencia del vuelo en el tamaño de los intrones, aunque los mecanismos que producen estas diferencias sigan aun siendo desconocidos (Zhang & Edwards 2012).

Como se puede ver en la Figura R1, la longitud del intrón 2 de los genes MHC de clase I de las aves del género *Serinus* es mayor que la del intrón 2 del pollo, por lo que no seguiría a priori la teoría comentada anteriormente sobre la reducción del tamaño de los intrones en aves voladoras. Sin embargo, para poder clarificar mejor este aspecto, se deberían de ampliar el número de intrones de estudio dentro de cada gen y ampliar el número de especies, contando con otros géneros de familias diferentes. También hay que recalcar que los genes de histocompatibilidad tienen un proceso evolutivo diferente y muy particular comparado con otro tipo de genes. Ninguno de los estudios que se han nombrado con anterioridad se basaba en intrones de genes de histocompatibilidad, por lo que faltan estudios en la literatura con los que comparar los resultados obtenidos y así poder obtener conclusiones.

### ***2.2 Análisis del porcentaje de similitud entre las secuencias del intrón 2***

Para conocer el grado de conservación entre las secuencias del intrón 2 de los genes MHC de clase I de las aves de estudio se calculó el porcentaje de similitud existente entre dichas secuencias, el cual se puede ver con detalle en la Tabla R4. Este análisis se complementa a su vez con el cálculo de las distancias genéticas entre las mismas secuencias del intrón 2 que se puede ver en la Tabla R3. La similitud entre las secuencias también se puede observar en la Figura R2, mediante un árbol filogenético basado en el método *Neighbor-joining*. Al realizar tres análisis independientes con las mismas secuencias de estudio, se pueden obtener resultados más fiables y una visión global del grado de conservación entre las secuencias del intrón 2 de las aves del género *Serinus*.

En general, las secuencias del intrón 2 de los genes MHC de clase I de las aves pertenecientes al género *Serinus* tienen un porcentaje de similitud bastante elevado, rondando en casi todos los casos el 98% (ver Tabla R4). Lo que indica que estas secuencias han estado sometidas a presiones selectivas que las han conservado, lo que puede deberse a que estas secuencias juegan un papel importante en la estabilidad y frecuencia del Sistema Principal de Histocompatibilidad del grupo de los canarios. Se da el caso incluso de dos secuencias que son idénticas en cada uno de sus nucleótidos, que son las pertenecientes al canario de cabeza estriada (*Serinus gularis*) y al canario de Brimstone (*Serinus sulphuratus*) (ver Tabla R4). Estas dos especies de canarios pertenecen al grupo de los canarios grandes africanos y aunque filogenéticamente están relacionados como se puede ver en las Figuras R4, R5 y D1, morfológicamente son



diferentes. Ambos son grandes, de unos 15 centímetros de longitud, pero mientras el canario de cabeza estriada es de un color blanquecino o gris, el canario de Brimstone es más verdoso y amarillo. El hecho de tener la misma secuencia del intrón 2 puede deberse entre otros muchos factores a que compartan hábitat, ya que a ambos se les puede localizar en regiones del sur de África, además de que filogenéticamente están muy relacionados.

Un aspecto a destacar de los resultados obtenidos es el hecho de que el intrón 2 del canario silvestre (*Serinus canaria*) se parezca tan poco al resto de intrones del grupo *Serinus*, dando como valor más alto de porcentaje de similitud un 34,63% (ver Tabla R4). Esta diferencia también se puede apreciar en el cálculo de las distancias genéticas (ver Tabla R3), donde el intrón 2 del canario silvestre tiene de media una distancia genética entorno al 0,6 (teniendo en cuenta que el valor máximo de distancia es 1, lo que implica que una secuencia es 100% diferente a otra), mientras que el resto de canarios muestran valores que están en torno al 0,03. Estas diferencias se pueden deber a la influencia en los patógenos existentes en las Islas Canarias, donde habita principalmente este ave. Aunque también puede deberse a otros procesos como la influencia del ambiente insular. Arnaiz-Villena y colaboradores (2007) realizaron un estudio de las moléculas de MHC de clase I en especies del género *Serinus*, donde también observaron grandes diferencias entre el canario silvestre y el resto de las aves del mismo género. Por lo que las fuerzas que afectan a la secuencia del intrón 2, no se centran solamente en esta zona, sino que afectan a toda la molécula de histocompatibilidad.

Otra especie que hay que destacar es el canario de pecho limón (*Serinus citrinipectus*). Este canario no presenta los valores tan elevados de similitud que en general se aprecian entre el resto de los canarios, (a excepción del canario silvestre, como se ha comentado anteriormente), ya que su porcentaje de similitud medio con respecto al resto de canarios es del 77% (ver Tabla R4). Esta ave pertenece al grupo de los canarios pequeños africanos (ver Figuras R4, R5 y D1) y habita principalmente en regiones del sur de África. Sin embargo el porcentaje de similitud de su secuencia no varía cuando se compara entre los canarios que pertenecen a su mismo grupo, es decir, los canarios pequeños africanos y el resto de los grupos. Las fuerzas de selección que hacen a este canario diferente en su secuencia del intrón 2 con respecto al resto son aún desconocidas con los datos de los que disponemos actualmente.



Por último, el estudio del porcentaje de similitud, muestra como las especies más diferentes al grupo de los canarios son el pinzón vulgar y el pollo, siendo esta última la más diferente (ver Tabla R4). Estas diferencias son de esperar ya el pollo pertenece a un orden, familia y género diferentes (Orden *Galliformes*, familia: *Phasianidae* y género *Gallus*) comparado con el grupo de los canarios (Orden *Passeriformes*, familia: *Fringillidae* y género *Serinus*). El pinzón vulgar pertenece al mismo orden y a la misma familia que el grupo de los canarios, aunque pertenece a un género distinto. Estas dos especies han evolucionado de manera diferente a lo largo de millones de años, por lo que es de esperar que sus moléculas de histocompatibilidad y con ello sus intrones sean diferentes al grupo de los *Serinus*.

En la literatura se encuentran pocos estudios que analizan la similitud entre las secuencias intrónicas de diferentes especies y géneros. Aruscavage y Bass (2000) estudiaron la enzima adenosindesaminasa, la cual está implicada en la edición del ARN, al cambiar las adenosinas por inosinas dentro de la doble cadena. Mediante un análisis filogenético compararon la secuencia de esta enzima en varias especies. Observaron un 94% de conservación en la secuencia del intrón 13 entre el ser humano y el pollo. Tras esta observación concluyeron que este intrón es funcionalmente importante (Aruscavage & Bass 2000). Por lo general se asume que las secuencias de los intrones a diferencia de la de los exones, no se conservan. Sin embargo poco a poco se van dando evidencias que respaldan la conservación de las secuencias intrónicas, aunque es difícil hacer conclusiones generales debido a que se han secuenciado pocos intrones de organismos diferentes.

Por su parte, Hare y Palumbi (2003), compararon mediante un método Bayesiano de alineamiento local 10 kb de secuencias intrónicas de 9 genes entre el ser humano, ballenas y focas. Como resultado de esa comparación, obtuvieron un 47% de conservación al comparar las especies por pares y un 28% al comparar los 3 taxones. En general, vieron que los segmentos de la secuencia del intrón más cercanos a los exones flanqueantes estaban significativamente más conservados que la secuencia del interior del intrón (Hare & Palumbi 2003). El nivel de conservación observado por estos autores superaba el nivel esperado por simple azar, por lo que se sugiere que las secuencias intrónicas están desempeñando un papel funcional más importante en la regulación de genes que lo que se había observado en estudios anteriores.

Rahman y colaboradores (2004), estudiaron el gen nPTB, el cual da lugar a un miembro de la familia de proteínas de unión de polipirimidinas que participan en el

procesamiento alternativo del pre-ARN mensajero. Estas proteínas se expresan principalmente en células neuronales. El gen nPTB tiene varias isoformas que se dan por el proceso de corte y empalme alternativo (*splicing* alternativo). Una de estas isoformas que sólo se detecta en tejido cerebral incluye la inserción del exón 10. La exclusión de este exón da lugar a otras isoformas. Al analizar este gen, encontraron que 297 nucleótidos intrónicos consecutivos que flanqueaban al exón 10 eran idénticos entre el ser humano, el ratón, el cerdo, la rata y el mono, mientras que 207 nucleótidos intrónicos consecutivos eran idénticos en el ADN de humanos y aves. También observaron una región de 2 kb dentro de un intrón que estaba conservada en mamíferos en un 85% y en aves en un 70%. Concluyeron que la notable conservación de este gran segmento de secuencia intrónica flanqueante sugeriría un papel importante en la selección del sitio de corte y empalme específico de tejido y podría funcionar en la regulación de la producción de nPTB funcional (Rahman et al. 2004).

El alto grado de conservación de las secuencias de estudio de este proyecto de tesis podría indicar que están desempeñando un papel importante en la regulación de las moléculas del Sistema Principal de Histocompatibilidad. Sin embargo, aunque la importancia de estas secuencias está clara, no se pueden hacer conclusiones firmes debido a la falta de estudios sobre la evolución de los intrones en genes de histocompatibilidad y a la escasez de secuencias intrónicas secuenciadas de genes en general. Para poder establecer conclusiones firmes serían necesarios más estudios sobre la evolución de estas secuencias, con un número mayor de intrones y con varios taxones de diferentes géneros y familias implicados.

### ***2.3 Análisis de las frecuencias nucleotídicas de las secuencias del intrón 2***

Para conocer la composición nucleotídica de las secuencias de estudio se calcularon las frecuencias de los cuatro tipos de bases que componen el ADN. Además se calculó el porcentaje de G+C, debido a su importancia. Las secuencias del intrón 2 de los genes MHC de clase I de las aves del género *Serinus*, muestran unos valores de porcentaje de G+C muy similares entre ellas (ver Tabla R2). El ave que posee un porcentaje mayor de G+C en su secuencia intrónica es el pollo, con un 78,6%, lo que supone un 20% más que el resto de especies. Este aumento es producido por el alto porcentaje de citosina que muestra esta especie (ver Tablas R1 y R2).

Los animales vertebrados de sangre caliente poseen regiones de hasta 300 kb ricas en nucleótidos de guanina y citosina (G+C), además estas regiones son más

abundantes en los genomas de las aves que en los de mamíferos (Bernardi et al. 1997). Varios estudios han descrito a las regiones ricas en G+C como las más estables frente a la desnaturalización térmica, lo que representa una ventaja para las aves, ya que tienen temperaturas y tasas metabólicas más elevadas (Bernardi & Bernardi 1986). El alto porcentaje de G+C del pollo con respecto al de las aves silvestres de canto del género *Serinus*, podría indicar que este ave posee tasas metabólicas más altas, aunque sería necesario analizar el porcentaje de G+C en más intrones para poder establecer conclusiones más firmes.

### **2.4 La importancia del estudio de los intrones**

En la última década el estudio de los intrones ha ido tomando cada vez más importancia. Ya no sólo se refieren a estos fragmentos como “ADN basura” sino que cada vez se va viendo cómo van tomando un papel más importante en la expresión de los genes y como pueden ser útiles a la hora de realizar análisis filogenéticos o como marcadores moleculares.

Chojnowski y colaboradores (2008) compararon la utilidad de las secuencias intrónicas y exónicas de dos genes de clatrina (un receptor de lipoproteínas), en la filogenia para intentar aclarar el origen de las Neoaves, grupo del cual forman parte aproximadamente el 95% de las especies de aves actuales. Lo que vieron fue que los intrones proporcionaban sitios más informativos, estaban sujetos a menos homoplasia y brindaban un mejor soporte para los clados que ya están bien aceptados por la comunidad científica. Por lo que estos investigadores sugerían que la evolución de los intrones sería más adecuada para determinar eventos de ramificación espaciados entre sí, como los que se dieron en el origen de las Neoaves (Chojnowski et al. 2008).

Anteriormente Fain y Houde (2004), ya habían empleado los intrones para estudiar la filogenia de las Neoaves. Más concretamente el intrón 7 del gen beta-fibrinógeno. Con el análisis de la filogenia basada en este intrón postularon que las Neoaves estaban formadas por dos clados hermanos con paralelismos ecológicos comparables a los encontrados entre los mamíferos marsupiales y placentarios (Fain & Houde 2004).

Otros autores como Jacobsen y colaboradores (2010), utilizaron los intrones para realizar análisis filogenéticos ante la gran cantidad de análisis existentes basados en un solo locus, el ADN mitocondrial. Vieron la necesidad de encontrar nuevos enfoques para el estudio de la evolución de las especies, por lo que optaron por realizar una

filogenia usando métodos tradicionales, pero basada en la combinación de 6 intrones independientes del cromosoma Z de un grupo de especies de oropéndolas (género *Icterus*). Lo que obtuvieron con este nuevo enfoque fue una filogenia de especies de oropéndolas en la cual se establecían los principales linajes y varios subclados de estas aves. Estas agrupaciones estaban apoyadas por publicaciones anteriores realizadas con ADN mitocondrial, lo que daba más soporte a esta nueva metodología de análisis filogenético (Jacobsen et al. 2010).

Otro uso que se le está dando a las secuencias intrónicas es el empleo de las mismas como marcadores moleculares. Pacheco y colaboradores (2002) utilizaron los intrones para investigar a dos poblaciones de aves marinas que parecían estar disminuyendo. Estos investigadores querían averiguar si se estaba produciendo una hibridación interespecífica utilizando los intrones como marcadores moleculares y así probar su utilidad frente a los marcadores morfológicos. Para ello analizaron la variación de cinco intrones nucleares y el gen del citocromo b mitocondrial en el mérgulo jaspeado (*Brachyramphus marmoratus*) y en el mérgulo piquicorto (*Brachyramphus brevirostris*). Sus resultados sugirieron que no había apenas hibridación entre estas dos especies y que habían sido independientes filogenéticamente desde hace unos 2 millones de años. Con ello concluyeron que los intrones eran una fuente importante de marcadores para analizar hibridación e introgresión de especies en peligro o en declive (Pacheco et al. 2002).

Actualmente se siguen valorando los intrones como posibles marcadores moleculares. Así, Gomes y colaboradores (2018), son investigadores que estudian la criptococosis, una de las micosis sistémicas más importantes en el mundo. Esta micosis es causada por diferentes genotipos de *Cryptococcus neoformans* y *Cryptococcus gattii*. Lo que hace diferente a estas especies es su ecología, epidemiología y susceptibilidad antifúngica. Para conocer su patogenicidad y susceptibilidad a diferentes medicamentos, utilizaron unos intrones mitocondriales, concretamente los del grupo I, ya que son polimórficos y su proceso de corte y empalme es inhibido por algunos fármacos. Lo que observaron fue que la presencia de estos intrones se asociaba estadísticamente con los genotipos menos patogénicos. Además vieron al realizar pruebas antimicóticas in vitro, que la presencia de estos intrones podría influir en la concentración inhibitoria mínima (MIC) del antifúngico anfotericina B y 5-fluorocitosina (Gomes et al. 2018).

Aunque hay que ampliar la investigación del uso de intrones como marcadores moleculares, estos estudios ponen de manifiesto la importancia de estas secuencias que siempre han estado relativamente olvidadas para la comunidad científica.

# CONCLUSIONES



En este proyecto de tesis se han obtenido datos moleculares procedentes de especies de aves silvestres pertenecientes al género *Serinus*. Se ha secuenciado la región correspondiente al intrón 2 de los genes MHC de clase I de 11 especies pertenecientes a este género. Además se han utilizado las secuencias del citocromo b mitocondrial de 19 especies del género *Serinus* para estudiar su clasificación y parentesco filogenético y su evolución en el tiempo; así se permite establecer las relaciones de antigüedad y proximidad genética entre las diferentes especies que forman este género.

Las conclusiones son:

- 1- El tamaño de la secuencia del intrón 2 de los genes MHC de clase I de las aves de canto pertenecientes al género *Serinus* es mayor que el de las secuencias del intrón 2 del pollo (*Gallus gallus*). Por tanto, el tamaño del intrón 2 de nuestro estudio que cuenta con 80 nucleótidos más que el intrón 2 del pollo, no permite establecer que tengan un MHC “mínimo esencial”, ni que este concepto sea universal en las aves.
- 2- Existe una conservación de la secuencia del intrón 2 de los genes MHC de clase I de las aves de canto del género *Serinus* de al menos 9 millones de años, lo que sugiere que esta secuencia juega un papel importante en la expresión de los genes de Histocompatibilidad y la existencia de una presión evolutiva positiva indeterminada sobre este intrón.
- 3- El intrón 2 de los genes MHC de clase I ha evolucionado de forma diferente al gen del citocromo b mitocondrial, como reflejan los análisis filogenéticos realizados con estos dos tipos de secuencias, como era de esperar por el funcionalismo conocido de ambas moléculas.
- 4- Mediante el análisis filogenético de las secuencias del citocromo b mitocondrial se ha realizado un estudio de la evolución de dicho gen en las aves de canto del género *Serinus*, lo que da una idea de sus tiempos de parentesco genético y especiación a lo largo de millones de años desde su aparición en la Tierra.





# BIBLIOGRAFÍA



- Abbas A, Lichtman A, Pober J (1994) *Cellular and molecular immunology*. Philadelphia, PA: WB Saunders.
- Afanassieff M et al (2001) At least one class I gene in restriction fragment pattern-Y (Rfp-Y), the second MHC gene cluster in the chicken, is transcribed, polymorphic and shows divergent specialization in antigen binding region. *J Immunol* 166:3324-3333.
- Akua T, Berezin I, Shaul O (2010) The leader intron of AtMHX can elicit, in the absence of splicing, low-level intron-mediated enhancement that depends on the internal intron sequence. *BMC.Plant.Biol.* 10:93.
- Anderson S et al (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
- Antoniou M et al (1998) Efficient 3'-end formation of human beta-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res* 26:721-729.
- Ardehali M, Lis J (2009) Tracking rates of transcription and splicing in vivo. *Nat Struct Mol Biol* 16:1123-1124.
- Arnaiz-Villena A, Areces C, Ruiz-del-Valle V (2012) El origen de los canarios. *Ornitología práctica* 53:3-11.
- Arnaiz-Villena A et al (2007) Evolution of the major histocompatibility complex class I genes in *Serinus canaria* from the Canary Islands is different from that of Asian and African continental *Serinus* species. *Journal of Ornithology* 148 (suppl.):S479-S484.
- Arnaiz-Villena A et al (1999) Rapid radiation of canaries (Genus *Serinus*). *Mol.Biol.Evol.* 16:2-11.
- Atkinson C, van Riper C (1991) Pathogenicity and epizootiology of avian haematozoa: Plasmodium, Leucocytozoon and Haemoproteus. In Loye J, Zuk M (eds) Bird-parasite interactions. New York: Oxford University Press.
- Aruscavage PJ, Bass BL (2000) A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. *RNA* 6:257-269.
- Bacon LD, Witter R (1994) An influence of the MHC on efficacy of Marek's disease vaccines. *Proc.5th World Congress on Genetics Applied to Livestock Production* 20:261-264.
- Barber J (1984) Has the Mangano-Protein of the Water Splitting reaction of Photosynthesis been isolated. *Trends Biochem Sci* 9:79-80.
- Barber LD, Parham P (1993) Peptide binding to major histocompatibility complex molecules. *Annu.Rev.Cell Biol* 9:163-206.
- Beck S et al (1996) Evolutionary dynamics of non-coding sequences within the class II region of the human MHC. *J Mol Biol* 255:1-13.

- Bensch S, Akesson S (2003) Temporal and spatial variation of hematozoans in Scandinavian willow warblers. *J Parasitol* 89:388-391.
- Bensch S et al (2000) Host specificity in avian blood parasites: a study of Plasmodium and Haemoproteus mitochondrial DNA amplified from birds. *Proc R Soc Lond B* 267:1583-1589.
- Berger R et al (1979) Regional mapping of the HLA on the short arm of chromosome 6. *Clin.Genet.* 15:245-251.
- Bernardi G, Bernardi G (1986) Compositional constraints and genome evolution. *J Mol Evol* 24:1-11.
- Bernardi G, Hughes S, Mouchiroud D (1997) The major compositional transitions in the vertebrate genome. *J Mol Evol* 44:44-51.
- Bjorkman PJ et al (1987a) Structure of the human class I histocompatibility antigen, HLA-A 2. *Nature* 329:506-512.
- Bjorkman PJ et al (1987b) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329:512-518.
- Bodmer W (1972) Evolutionary significance of the HLA system. *Nature* 237:139-145.
- Bollmer J, Vargas F, Parker P (2007) Low MHC variation in the endangered Galapagos penguin (*Spheniscus mendiculus*). *Immunogenetics* 59:593-602.
- Bonneaud C et al (2006) MHC alleles confer local resistance to malaria in a wild passerine. *Evolution* 60:383-389.
- Bonneaud C et al (2004) Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics* 55:855-865.
- Bornstein P et al (1988) Interactions between the promoter and first intron are involved in transcriptional control of alpha 1 (I) collagen gene expression. *Mol Cell Biol* 8:4851-4857.
- Briles W et al (1993) A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics* 37:408-414.
- Briles W, McGibbon W, Irwin DM (1950) On multiple alleles affecting cellular antigens in the chicken. *Genetics* 35:633-640.
- Briles W, Briles R (1980) A search for "duplicate" recombinants between B-F and B-G regions in the chicken B complex. *Anim.Blood Groups Biochem.Genet.* 11:38-39.
- Briles W, Briles R (1982) Identification of haplotypes of the chicken major histocompatibility complex (B). *Immunogenetics* 15:449-459.
- Briles W, McGibbon W (1948) Heterozygosity of inbred lines of chickens of two loci effecting cellular antigens. *Genetics* 33:605.

- Brocke P et al (2002) HLA-DM, HLA-DO and tapasin: functional similarities and differences. *Curr.Opin.Immunol.* 14:22-29.
- Brown T. 2006. *Genomes*. 3rd ed. New York: Garland Science.
- Brugge J, Erikson R (1977) Identification of a transformation-specific antigen by an avian sarcoma virus. *Nature* 269:346-348.
- Buchman A, Berg P (1988) Comparison of intron-dependent and intron-independent gene expression. *Mol.Cell.Biol.* 8:4395-4405.
- Castillo-Davis et al (2002) Selection for shorts introns in highly expressed genes. *Nature*. 31:415-418.
- Cepellini R et al (1967) Genetics of leukocyte antigens. A family study of segregation and linkage. In Curtini ES, Mattiuz PL, Tosi RM (eds) *Histocompatibility Testing 1967*. Copenhagen: Munksgaard.
- Chappell P et al (2015) Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *eLife* 4:e05345.
- Chojnowski JL, Kimball RT, Braun EL (2008) Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene* 410:89-96.
- Chorev M, Carmel L (2012) The function of introns. *Frontiers in Genetics* 3:1-15.
- Clement P, Harris P, Davies J. 1993. *Finches and Sparrows*. London: Princeton University Press.
- Collins W et al (1977) The B locus (MHC) in the chicken: Association with the fate of RSV-induced tumors. *Immunogenetics* 5:333-343.
- Connell S et al (2012) Avian resistance to *Campylobacter jejuni* colonization is associated with an intestinal immunogene expression signature identified by mRNA sequencing. *PLoS ONE* 7:e40409.
- Dausset J (1958) Iso-leuko-antibodies. *Acta Haematol.* 20:156-166.
- Dausset J, Ivanyi P, Ivanyi D. 1965. *Tissue alloantigens in humans: identification of a complex system (Hu-1)*. In Dausset J et al (eds) *Histocompatibility Testing 1965*. Copenhagen: Munksgaard.
- Denzin LK et al (1997) Negative regulation by HLA-DO of MHC class II-restricted antigen processing. *Science* 278:106-109.
- Doherty PC, Zinkernagel RM (1975) Enhanced immunologic surveillance in mice heterozygous at the H2 complex. *Nature* 256:50-52.
- Dredge B, Darnell R (2003) Nova regulates GABA(A) receptor gamma 2 alternative splicing via a distal downstream UCAU rich intronic splicing enhancer. *Mol Cell Biol* 23:4687-4700.

- Dvir H et al (2010) Structural basis for lipid-antigen recognition in avian immunity. *J Immunol* 184:2504-2511.
- Edwards S, Hedrick P (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol Evol* 13:305-311.
- Eklblom R, Grahn M, Hoglund J (2003) Patterns of polymorphism in the MHC class II of a non-passerine bird, the great snipe (*Gallinago media*). *Immunogenetics* 54:734-741.
- Encyclopaedia Britannica. 1999. Sutton, Surney: Encyclopaedia Britannica International Ltd.
- Fain MG, Houde P (2004) Parallel radiations in the primary clades of birds. *Evolution* 58:2558-2573.
- Falk K et al (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351:290-296.
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J.Mol.Evol.* 17:368-376.
- Gasper JS et al (2001) Songbird genomics: analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (*Agelaius phoeniceus*). *Genomics* 75:26-34.
- Gaunt A, Wells M (1973) Models of syringeal mechanisms. *Am Zool* 13:1227-1247.
- Geng T, Guan X, Smith E (2015) Screening for genes involved in antibody response to sheep red blood cells in the chicken, *Gallus gallus*. *Poult.Sci.* 94:2099-2107.
- Ghosh P et al (1995) The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. *Nature* 378:457-462.
- Gilson P et al (2006) Complete nucleotide sequence of the chlorarachniophyte nucleophorm: nature's smallest nucleus. *Proc Natl Acad Sci U.S.A* 103:9566-9571.
- Glynne R et al (1991) A proteasome-related gene between the two ABC transporter loci in the class II region of the human MHC. *Nature* 353:357-360.
- Gomes FE et al (2018) Polymorphism in mitochondrial group I introns among *Cryptococcus neoformans* and *Cryptococcus gattii* genotypes and its association with drug susceptibility. *Front Microbiol* 9:86.
- Gomez-Prieto P et al (2010) HLA-G, -F and -E: Polymorphism, Function, and Evolution. In Mehra NK (ed) *The HLA Complex in Biology and Medicine. A Resource Book*, First ed. New Delhi: Jaypee Brothers Medical Publishers.
- Gorer PA (1936) The detection of a hereditary antigenic difference in the blood of mice by means of human group A serum. *Journal of Genetics* 32:17-31.

- Goto R et al (2009) BG1 has a major role in MHC-linked resistance to malignant lymphoma in the chicken. *Proc Natl Acad Sci U.S.A* 106:16740-16745.
- Graur D, Li W. 2000. *Fundamentals of molecular evolution*. Sunderland, Massachusetts: Sinauer Associates, inc.
- Greenewalt C (1969) How birds sing. *Sci Am* 221:126-139.
- Grey HM et al (1973) The small subunit of HL-A antigens is beta 2-microglobulin. *J Exp.Med* 138:1608-1612.
- Gruss P et al (1979) Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40. *Proc Natl Acad Sci U.S.A* 76:4317-4321.
- Gubb D (1986) Intron-delay and the precision of expression of homoeotic gene products in *Drosophila*. *Dev Genet* 7:119-131.
- Guillemot F, Auffray C (1989) A molecular map of the chicken B complex. *Prog.Clin.Biol.Res.* 307:169-176.
- Guillemot F et al (1988) A molecular map of the chicken major histocompatibility complex: the class II beta genes are closely linked to the class I genes and the nucleolar organizer. *EMBO J* 7:2775-2785.
- Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95-98.
- Hansen M, Van Zandy J, Law G (1967) Differences in susceptibility to Marek's disease in chickens carrying two different B locus blood group alleles. *Poult.Sci.* 46:1268.
- Hare MP, Palumbi SR (2003) High intron sequence conservation across three mammalian orders suggests functional constraints. *Mol Biol Evol* 20:969-978.
- Hartmann B, Valcarcel J (2009) Decrypting the genome's alternative messages. *Curr Opin Cell Biol* 21:377-386.
- Hedrick P (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution* 56:1902-1908.
- Hee C et al (2010) Structure of a classical MHC class I molecule that binds "non-classical" ligands. *PLoS Biol* (8:e1000557).
- Hee C et al (2009) Expression, purification and preliminary X-ray crystallographic analysis of the chicken MHC class I molecule YF1\*7.1. *Acta Crystallogr* F65:422-425.
- Hess C, Edwards S (2002) The evolution of the major histocompatibility complex in birds. *Bioscience* 52:423-431.



- Hofmann A et al (2003) v-src oncogene-specific carboxy-terminal peptide is immunoprotective against Rous sarcoma growth in chickens with MHC class I allele B-F12. *Vaccine* 21:4694-4699.
- Horton R et al (2008) Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics* 60:1-18.
- Horton R et al (2004) Gene map of the extended human MHC. *Nat.Rev.Genet.* 5:889-899.
- Howell N, Gilbert K (1988) Mutational analysis of the mouse mitochondrial cytochrome -b gene. *J Mol Biol* 203:607-618.
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755.
- Hughes AL, Hughes MK (1995) Small genomes for better flyers. *Nature.* 377:391.
- Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet* 32:415-435.
- Humphreys T, Reinherz EL (1994) Invertebrate immune recognition, natural immunity and the evolution of positive selection. *Immunol.Today* 15:316-320.
- Hunt H et al (2006) At least one YMHCI molecule in the chicken is alloimmunogenetic and dinamically expressed on spleen cells during development. *Immunogenetics* 58:297-307.
- Iizuka K et al (2003) Genetically linked C-Type lectin-related ligands for the NKRP1 family of natural killer cell receptors. *Nat.Immunol.* 4:801-807.
- Jacob J et al (2000) The major and a minor class II beta-chain (B-LB) gene flank the Tapasin gene in the B-F/B-L region of the chicken major histocompatibility complex. *Immunogenetics* 51:138-147.
- Jacobsen F, Friedman NR, Omland KE (2010) Congruence between nuclear and mitochondrial DNA: Combination of multiple nuclear introns resolves a well-supported phylogeny of New World orioles (*Icterus*). *Mol Phylogenet Evol* 56:419-427.
- Jarvi S et al (2004) Natural selection of the major histocompatibility complex (MHC) in Hawaiian honeycreepers (Drepanidinae). *Mol Ecol* 13:2157-2168.
- Jensen CO (1903) Experimentelle untersuchungen iiber krebs bei mausen. *Zentralbl Bakteriol Parasitenkd Infektkr* 34:28.
- Jordan BR et al (1985) HLA class I genes: from structure to expression, serology and function. *Immunol.Rev.* 84:73-92.
- Juneau K et al (2006) Introns regulate RNA and protein abundance in yeast. *Genetics* 174:511-518.

- Katunuma N, Matsunaga Y, Saibara T (1994) Mechanism and regulation of antigen processing by cathepsin B. *Adv Enzyme Regul.* 34:145-158.
- Kaufman J et al (1999) Gene organisation determines evolution of function in the chicken MHC. *Immunol.Rev* 167:101-117.
- Kaufman J et al (1990) Size polymorphism of chicken major histocompatibility complex encoded B-G molecules is due to length variation in the cytoplasmatic heptad repeat region. *Proc Natl Acad Sci U.S.A* 87:8277-8281.
- Kaufman J, Volk H, Wallny HJ (1995) A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol.Rev* 143:63-88.
- Khodor Y et al (2011) Nascent-seq indicates widespread cotranscriptional pre-RNA splicing in *Drosophila*. *Genes Dev* 25:2502-2512.
- Kjer-Nielsen L et al (2012) MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature* 491:717-723.
- Klein J. 1986. *Natural history of the Major Histocompatibility Complex*. New York, USA: J. Willey and Sons.
- Koch M et al (2007) Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity.* 27:885-899.
- Koopmann JO, Hammerling GJ, Momburg F (1997) Generation, intracellular transport and loading of peptides associated with MHC class I molecules. *Curr.Opin.Immunol.* 9:80-88.
- Kumar S, Stecher G, Tamura K (2016) MEGA 7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870-1874.
- Le Hir H, Nott A, Moore M (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 28:215-220.
- LePage K et al (2000) Rfp-Y genotype affects the fate of Rous sarcomas in B2B5 chickens. *Immunogenetics* 51:751-754.
- Li L, Bouvier M (2004) Structures of HLA-A\*1101 complexed with immunodominant nonamer and decamer HIV-1 epitopes clearly reveal the presence of a middle, secondary anchor residue. *J Immunol.* 172:6175-6184.
- Little CC, Johnson BW (1922) The inheritance of susceptibility to implants of splenic tissue in mice. *Proceedings of the Society for Experimental Biology and Medicine* 19:163.
- Loeb L (1908) Ueber Entstehung eines Sarkoms nach Transplantation eines Adenocarcinoms einer japanischen Maus. *Journal of Cancer Research and Clinical Oncology* 7:80.

- Lopez-Bigas N et al (2005) Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579:1900-1903.
- Lowy E (2003) Evolución y sistema principal de histocompatibilidad de canarios (género *Serinus*), Universidad Complutense Madrid, Director: Dr. A. Arnaiz-Villena, Sobresaliente *Cum Laude*.
- Luehrsen K, Walbot V (1991) Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. *Mol.Gen.Genet.* 225:91-93.
- Lynch M (2002) Intron evolution as a population-genetic process. *PNAS* 99:6118-6123.
- Lynch M. 2007. *The Origins of Genome Architecture*. 1st Edn ed. Sunderland: Sinauer Associates Inc.
- Maccubbin D, Schierman L (1986) MHC restricted cytotoxic response of chicken T cells: expression, augmentation, and clonal characterization. *J Immunol* 136:12-16.
- Madden DR (1995) The three-dimensional structure of peptide-MHC complexes. *Annu.Rev.Immunol.* 13:587-622.
- Mair G et al (2000) A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* 6:163-169.
- Malissen M, Malissen B, Jordan BR (1982) Exon/intron organization and complete nucleotide sequence of an HLA gene. *Proc.Natl.Acad.Sci.U.S.A* 79:893-897.
- Mascarenhas D et al (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant.Mol.Biol* 15:913-920.
- Miller H, Lambert D (2004) Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* 56:178-191.
- Miller MM et al (1994) Two MHC class I and two MHC class II genes map to the chicken Rfp-Y system outside the B complex. *Proc Natl Acad Sci U.S.A* 91:4397-4401.
- Miller MM et al (1996) Assignment of Rfp-Y to the chicken major histocompatibility complex/NOR microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region. *Proc Natl Acad Sci U.S.A* 93:3958-3962.
- Miller M et al (1991) Immunoglobulin variable-region-like domains of diverse sequence within the major histocompatibility complex of the chicken. *Proc Natl Acad Sci U.S.A* 88:4377-4381.
- Miller M et al (1990) Antigens similar to major histocompatibility complex B-G are expressed in the intestinal epithelium in the chicken. *Immunogenetics* 32:45-50.

- Miller M, Taylor R (2016) Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poult.Sci.* 95:375-392.
- Miller MM et al (2014) Mapping genes to chicken microchromosome 16 and discovery of olfactory and scavenger receptor genes near the major histocompatibility complex. *J Hered* 105:203-215.
- Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via polymerase catalyzed chain reaction. *Meth Enzimol* 155:335-350.
- Mungall AJ et al (2003) The DNA sequence and analysis of human chromosome 6. *Nature* 425 (6960):805-811.
- Nei M (1972) Genetic distances between populations. *Am.Nat.* 106:283-292.
- Nei M. 1987. *Molecular Evolutionary Genetics*. New York, NY (USA): Columbia University Press.
- Nilsen T, Graveley B (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463:457-463.
- Pacheco NM, Congdon BC, Friesen VL (2002) The utility of nuclear introns for investigating hybridization and genetic introgression: a case study involving *Brachyramphus* murrelets. *Conserv Genet* 3:175-182.
- Parham P et al (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. *PNAS* 85:4005-4009.
- Peterson PA, Rask L, Lindblom JB (1974) Highly purified papain-solubilized HL-A antigens contain beta2-microglobulin. *Proc.Natl.Acad.Sci.U.S.A* 71:35-39.
- Pharr G, Gwynn A, Bacon LD (1996) Histocompatibility antigen(s) linked to Rfp-Y (MHC-like) genes in the chicken. *Immunogenetics* 45:52-58.
- Pieters J (1997) MHC class II restricted antigen presentation. *Curr.Opin.Immunol.* 9:89-96.
- Pink JR et al (1977) A three-locus model for the chicken major histocompatibility complex. *Immunogenetics* 5:203-216.
- Plachy J, Benda V (1981) Location of the gene responsible for Rous sarcoma regression in the B-F region of the B complex (MHC) of the chicken. *Folia Biol.(Praha)* 27:363-368.
- Ploegh HL, Orr HT, Strominger JL (1981) Major histocompatibility antigens: the human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. *Cell* 24:287-299.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817-818.

- Rahman L et al (2004) Evolutionary conservation of a 2-kb intronic sequence flanking a tissue-specific alternative exon in the PTBP2 gene. *Genomics* 83:76-84.
- Rammensee HG (1995) Chemistry of peptides associated with MHC class I and class II molecules. *Curr.Opin.Immunol.* 7:85-96.
- Rammensee HG, Falk K, Rotzschke O (1993) Peptides naturally presented by MHC class I molecules. *Annu.Rev.Immunol.* 11:213-244.
- Richardson DS, Westerdahl H (2003) MHC diversity in two *Acrocephalus* species: the outbred Great reed warbler and the inbred Seychelles warbler. *Mol Ecol* 12:3523-3529.
- Ricklefs R, Fallon S (2002) Diversification and host switching in avian malaria parasites. *Proc R Soc Lond B* 269:885-892.
- Robinson J et al (2015) The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* 43 (Database issue):D423-D431.
- Rogers S et al (2003) Analysis of a part of the chicken Rfp-Y region reveals two novel lectin genes, the first complete genomic sequence of a class I alpha-chain gene, a truncated class II beta-chain gene, and a large CR1 repeat. *Immunogenetics* 55:100-108.
- Saiki R et al (1985) Enzymatic amplification of  $\alpha$ -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-1354.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* 4:406-425.
- Sakharkar M, Chow V, Kagueane P (2004) Distributions of exons and introns in the human genome. *In Silico Biol* 4:387-393.
- Salomonsen J et al (2014) Sequence of a complete chicken BG haplotype shows dynamic expansion and contraction of two genes lineages with particular expression patterns. *PLoS.Genet.* 10:e1004417.
- Salomonsen J et al (2003) The properties of the single chicken MHC classical class II alpha chain (B-LA) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. *Immunogenetics* 55:605-614.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Sambrook J, Fritsch EF, Maniatis T (eds). New York: Cold Spring Harbor Laboratory Press.
- Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U.S.A* 71:5463-5467.
- Saper MA, Bjorkman PJ, Wiley DC (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J Mol Biol* 219:277-319.

- Saraste M (1984) Location of heme-binding sites in the mitochondrial cytochrome-b. *FEBS Lett.* 166:367-372.
- Schat K, Taylor R, Briles W (1994) Resistance to Marek's disease in chickens with recombinant haplotypes to the major histocompatibility (B) complex. *Poult.Sci.* 73:502-508.
- Schierman L, Nordskog A (1961) Relationship of blood type to histocompatibility in chickens. *Science* 134:1008-1009.
- Schierman L, Watanabe D, McBride R (1977) Increased growth of Rous sarcomas in chickens pretreated with formalinized syngeneic tumor cells. *Eur.J.Immunol.* 7:710-713.
- Schmidt M et al (1999) Sequence information within proteasomal prosequences mediates efficient integration of beta-subunits into the 20 S proteasome complex. *J Mol Biol* 288:117-128.
- Schmucker D et al (2000) Drosophila Dscam in an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101:671-684.
- Schwartz S et al (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res* 18:88-103.
- Schohy S et al (2000) Identification of an enhancer and an alternative promoter in the first intron of the alpha-fetoprotein gene. *Nucleic Acids Res* 28:3743-3751.
- Shabalina SA et al (2010) Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. *Mol Biol Evol* 27:1745-1749.
- Shiina T et al (2007) Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious diseases. *J Immunol* 178:7162-7172.
- Sibley CG, Ahlquist J. 1990. *Phylogeny and classification of birds*. New Haven, Conn.: Yale University Press.
- Sibley C. 1995. *Birds of the World*. Thayer Birding Software.
- Simonson M, Arnul M, Sorensen P (1989) *The chicken MHC and its importance*. In: Improving genetic disease resistance in farm animals. Ed. by Van der Zijpp and Sybesma, Kluwer Academic Publishers, Dordrecht, The Netherlands, 42-62.
- Singh J, Padgett RA (2009) Rates of in situ transcription and splicing in large human genes. *Nat.Struct.Mol.Biol.* 16:1128-1133.
- Sloan VS et al (1995) Mediation by HLA-DM of dissociation of peptides from HLA-DR. *Nature* 375:802-806.
- Snell GD (1948) Methods for the study of histocompatibility genes. *Journal of Genetics* 49:87-108.



- Solinhac R et al (2010) Integrative mapping analysis of chicken microchromosome 16 organization. *BMC Genomics* 11:616.
- Spies T et al (1990) A gene in the human major histocompatibility complex class II region controlling the class I antigen presentation pathway. *Nature* 348:744-747.
- Spike C et al (1996) RFLP mapping of expressed sequence tags in the chicken. *J Hered* 87:6-9.
- Stern LJ et al (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368:215-221.
- Stewart CA et al (2004) Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res.* 14:1176-1187.
- Strand T et al (2007) The MHC class II of the black grouse (*Tetrao tetrix*) consists of low numbers of B and Y genes with variable diversity and expression. *Immunogenetics* 59:725-734.
- Swofford DL. 2002. *PAUP\* Phylogenetic Analysis Using Parsimony (\* and other methods)* version 4. Sunderland, Massachusetts: Sinauer Associates.
- Taylor R (2004) Major histocompatibility (B) complex control of responses against Rous sarcomas. *Poult.Sci.* 83:638-649.
- The MHC sequencing consortium (1999) Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 401:921-923.
- Thomsen M et al (1994) Recombination fractions in the HLA system based on the data set 'provinces Francaises': indications of haplotype-specific recombination rates. *Eur.J Immunogenet.* 21:33-43.
- Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol.Biol.Evol.* 15:1647-1657.
- Thummel C, Burtis K, Hogness D (1990) Spatial and temporal patterns of E74 transcription during *Drosophila* development. *Cell* 61:101-111.
- Tonnelle C, DeMars R, Long EO (1985) DO beta: a new beta chain gene in HLA-D with a distinct regulation of expression. *EMBO J* 4:2839-2847.
- Tourmente S et al (1993) Enhancer and silencer elements within the first intron mediate the transcriptional regulation of the beta 3 tubulin gene by 20-hydroxyecdysone in *Drosophila* Kc cells. *Insect Biochem.Mol.Biol.* 23:137-143.
- Trowsdale J et al (1990) Sequences encoded in the class II region of the MHC related to the 'ABC' superfamily of transporters. *Nature* 348:741-744.
- Trowsdale J, Kelly A (1985) The human HLA class II alpha chain gene DZ alpha is distinct from genes in the DP, DQ and DR subregions. *EMBO J* 4 (9):2231-2237.

- Trowsdale J, Ragoussis J, Campbell RD (1991) Map of the human MHC. *Immunol.Today* 12:443-446.
- Tulp A et al (1994) Isolation and characterization of the intracellular MHC class II compartment. *Nature* 369:120-126.
- Wahl M, Will C, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* 136:701-718.
- Waltari E y Edwards SV (2002) Evolutionary dynamics of intron size, genome size, and physiological correlates in Archosaurs. *Amer Naturalist* 160:539-552.
- Westerdahl H (2007) Passerine MHC: genetic variation and disease resistance in the wild. *J Ornithol* 148:S469-S477.
- Westerdahl H et al (2005) An association between malaria infection and MHC in great reed warblers. *Proc R Soc Lond B* 272:1511-1518.
- Westerdahl H et al (2004) MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. *Heredity* 92:534-542.
- Westerdahl H, Wittzell H, von Schantz T (1999) Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. *Immunogenetics* 49:158-170.
- Widger W et al (1984) Sequence homology and structural similarity between cytochrome-b of mitochondrial complex-III and the chloroplast-B6-F complex-Position of the Cytochrome-b hemes in the membrane. *Proc Natl Acad Sci U.S.A* 81:674-678.
- Winogradov A (1999) Intron-genome size relationship on a large evolutionary scale. *J Mol Evol.* 49:376-384.
- Wittzell H et al (1999) Concerted evolution of two MHC class II B loci in pheasants and domestic chickens. *Mol Biol Evol* 16:479-490.
- Wu C, Macleod I, Su AI (2013a) BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.* 41 (Database issue):D561-D565.
- Wu G et al (2015) Splenic gene expression profiling in White Leghorn layer inoculated with the Salmonella enterica serovar Enteritidis. *Anim.Genet.* 46:617-626.
- Wu JY et al (2013b) Systematic analysis of intron size and abundance parameters in diverse lineages. *Sci China Life Sci.* 56:968-974.
- Yoshimoto R et al (2009) Isolation and characterization of post splicing lariat-intron complexes. *Nucleic Acids Res* 37:891-902.
- Zajonc D et al (2008) The crystal structure of avian CD1 reveals a smaller, more primordial antigen binding pocket compared to mammalian CD1. *Proc Natl Acad Sci U.S.A* 105:17925-17930.



- Zamora J et al (2006) Conjoint mitochondrial phylogenetic trees for canaries (*Serinus* spp.) and goldfinches (*Carduelis* spp.) show several specific polytomies. *Ardeola* 53:1-17.
- Zhang Q, Edwards SV (2012) The evolution of intron size in Amniotes: A role for powered flight? *Genome Biol Evol* 4:1033-1043.
- Zinkernagel RM, Doherty PC (1979) MHC-restricted cytotoxic T cells: studies on the biological role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function, and responsiveness. *Adv.Immunol.* 27:51-177.

# ANEXOS



**ANEXO 1: ABREVIATURAS MÁS FRECUENTES**

|              |  |
|--------------|--|
| <b>ADN</b>   | Ácido desoxirribonucleico                        |
| <b>mtADN</b> | Ácido desoxirribonucleico mitocondrial           |
| <b>ARN</b>   | Ácido ribonucleico                               |
| <b>ARNm</b>  | Ácido ribonucleico mensajero                     |
| <b>ARNt</b>  | Ácido ribonucleico transferente                  |
| <b>snRNA</b> | Ácido ribonucleico pequeño nuclear               |
| <b>BI</b>    | Inferencia Bayesiana                             |
| <b>β2m</b>   | β2-microglobulina                                |
| <b>dNTPs</b> | Desoxirribonucleicos trifosfato                  |
| <b>HLA</b>   | Antígeno leucocitario humano                     |
| <b>l</b>     | Litro  |
| <b>M</b>     | Molar  |
| <b>ma</b>    | Millones de años                                 |
| <b>MHC</b>   | Sistema Principal de Histocompatibilidad         |
| <b>ML</b>    | Maximum Likelihood                               |
| <b>NJ</b>    | Neighbor-Joining                                 |
| <b>NK</b>    | Natural Killer                                   |
| <b>NOR</b>   | Región de organización nucleolar                 |
| <b>pb</b>    | pares de bases                                   |
| <b>PCR</b>   | Reacción en cadena de la polimerasa              |
| <b>μg</b>    | Microgramo                                       |
| <b>RFLP</b>  | Fragmento de restricción de longitud polimórfica |

## ANEXO 2: CURRICULUM VITAE

### Formación académica

- Grado en Ciencias Ambientales por la Universidad de Alcalá
- Máster en Microbiología y Parasitología por la Universidad Complutense
- Postgrado en Investigación Biomédica por la Universidad Complutense

### Publicaciones científicas

- 1.- Rey D, Amirzargar A, Areces C, Enriquez-de-Salamanca E, Marco J, Abd-El-Fatah-Khalil S, Fernandez-Honrado M, **Muñiz E**, Martin-Villa JM, Arnaiz-Villena A. *Gorgan (Turkmen in Iran) HLA genetics: transplantation, pharmacogenomics and anthropology*. Immunological Investigations, doi: 10.3109/08820139.2014.936938, 2014.
- 2.- Arnaiz-Villena A, Ruiz-del-Valle V, Gomez-Prieto P, Rey D, Enriquez-de-Salamanca E, Marco J, **Muñiz E**, Martin-Villa M, Areces C. *Carduelini New Systematics: Crimson-winged Finch (Rodhopenchys sanguineus) is included in "Arid-Zone" Carduelini Finches by Mitochondrial DNA Phylogeny*. The Open Ornithology Journal, 7:55-62, 2014.
- 3.- Arnaiz-Villena A, Areces C, Enriquez-de-Salamanca M, Abd-El-Fatah-Khalil S, Marco J, **Muñiz E**, Fernandez-Honrado M, Martin-Villa M, Rey D. *Pacific Islanders and Amerindian relatedness according to HLA autosomal genes*. Internacional Journal of Modern Anthropology, 7:44-67, 2014.
- 4.- Arnaiz-Villena A, **Muñiz E**, Palacio-Grüber J, Martinez-Laso J. *The Peopling of America: Siberians, Pacific Islanders and Ancient Trans-Atlantic Europeans according to HLA genetics and anthropology*. Nova Publishers. 2015
- 5.- Amirzargar A, Rey D, **Muñiz E**, Palacio-Grüber J, Nikbin B, Nicknam H, Khosravi F, Joshghan H, Areces C, Enriquez-de-Salamanca M, Martinez-Quiles N, Martín-Villa M, Arnaiz-Villena A. *Kurds HLA Genes: Its implications in Transplantation and Pharmacogenomics*. Open Medicine Journal, 2: 43-47. 2015
- 6.- Recio MJ, Martinez-Laso J, **Muñiz E**, Palacio-Grüber J, Rey D, Martín-Villa M, Martinez-Quiles N, Campos C, Arnaiz-Villena A. *Major Histocompatibility complex-DMB allelic diversity in old and new world nonhuman primates: Intraspecies pattern of evolution*. International Journal of Modern Anthropology. 8:25-47. 2015
- 7.- Arnaiz-Villena A, Enríquez-de-Salamanca M, **Muñiz E**, Vasile L, Carballo A, Martín-Villa M, Martinez-Quiles N, Areces C, Rey D. *Easter Island and Aleuts relatedness to Amerindians and Pacific Islanders according to HLA genes*. Tissue Antigens. 85: 396.DOI: 10.1111/tan.12557. 2015

- 8.- Arnaiz-Villena A, Vargas-Alarcón G, Rey D, Enriquez-de-Salamanca M, **Muñiz E**, Areces C, Martín-Villa M, Martínez-Quiles N, Cruz-Robles D, López-Pacheco F. *Mixtec Mexican Amerindians: an HLA allele study for America peopling, pharmacogenomics and transplantation*. Tissue Antigens. 85: 397.DOI: 10.1111/tan.12557. 2015
- 9.- Arnaiz-Villena A, Gómez-Prieto P, Enriquez-de-Salamanca M, Rey D, **Muñiz E**, Corrales A, Martín-Villa M, Martínez-Quiles N, Areces C. *MHC variability in wild Songbirds: normal intron size and molecule evolution suggest that a minimal essential MHC does not exist in birds*. Tissue Antigens. 85: 397.DOI: 10.1111/tan.12557. 2015
- 10.- Arnaiz-Villena A, Areces C, Enriquez-de-Salamanca M, Marco J, **Muñiz E**, Fernandez-Honrado M, Rey D. *Evidence of Paleohistorical Atlantic and Pacific transoceanic genetic and cultural contacts*. Human biology. Issue supplement S60, 156:1-334. 2015
- 11.- Arnaiz-Villena A, **Muñiz E**, Campos C, Gómez-Casado E, Tomasi S, Martínez-Quiles N, Martín-Villa JM, Palacio-Grüber J. *Origin of Ancient Canary Islanders (Guanches): presence of atlantic/Iberian HLA and Y chromosome genes and Ancient Iberian language*. International Journal of Modern Anthropology. 8:67-93. 2015.
- 12.- Arnaiz-Villena A, Rey D, Recio MJ, **Muñiz E**, Palacio-Grüber J, Campos C, Martínez-Quiles N, Martín-Villa JM, Martínez-Laso J. *HLA-DMB in Amerindians: specific association of DMB\*01:03:01/DRB1 alleles*. Human Immunology. doi: 10.1016/j.humimm.2016.02.011. 2016.
- 13.- Arnaiz-Villena A, Enriquez-de-Salamanca M, Palacio-Grüber J, Campos C, Camacho A, Martín-Villa JM, Martínez-Quiles N, **Muñiz E**. *Characterization and functional implications of the two new HLA-G alleles found in Amerindian and Caribbean populations*. Human Immunology, doi:10.1016/j.humimm.2016.01.006. 2016.
- 14.- Arnaiz Villena A, Cruz Robles D, Vargas Alarcon G, Lopez Pacheco F, Rodriguez Perez JM, **Muniz E**, Arriaga JC, Granados J, Escobedo G, Frago JM. *Novel description of CYP11B2 -344 T>C gene polymorphism related to hypertension in Mexican Amerindians: Teenek, Mixtec and Mayans*. International Journal of Modern Anthropology. 9:52-64. 2016.
- 15.- Arnaiz-Villena A, **Muñiz E**, Palacio-Grüber J, Campos C, Martín-Villa M, Gómez-Casado E, López-Pacheco F, Silvera C. *HLA genes in Chimila Amerindians (Colombia): transplantation, epidemiology, pharmacogenomics and anthropology*. International Journal of Modern Anthropology. 9: 91-116. 2016
- 16.- Arnaiz-Villena A, **Muñiz E**, Palacio-Grüber J, Campos C, Martín-Villa M, Gómez-Casado E, López-Pacheco F, Silvera C. *Ancestry of Amerindians and its impact in anthropology, transplantation, HLA pharmacogenomics and epidemiology by HLA study in Wiwa Colombian population*. Open Medicine Journal. 3. 2016.
- 17.- Arnaiz-Villena A, Palacio-Gruber J, **Muñiz E**, Campos C, Alonso-Rubio J, Gómez-Casado E, Salih S, Martín-Villa M, Al-Qadi R. *Genetic HLA Study of*

- Kurds in Iraq, Iran and Tbilisi (Caucasus, Georgia): Relatedness and Medical Implications*. DOI:10.1371/journal.pone.0169929. PlosOne. 2017.
- 18.- Arnaiz-Villena A, Carballo A, **Muñiz E**, Campos C, Tejedor B, Martín-Villa M, Palacio-Gruber J. *HLA genes in Atlantic Celtic populations: are Celts Iberians?* DOI: <http://dx.doi.org/10.4314/ijma.v1i10.2>. 2017.
  - 19.- Arnaiz-Villena A, Ruiz-del-Valle V, **Muñiz E**, Palacio-Gruber J, Campos C, Gómez-Casado E, Martín-Villa JM, Ignacio Serrano-Vela. *Major histocompatibility complex allele persistence in Eurasia and America genus Carduelis during million years*. *The Open Ornithology Journal*. 10: 92-104. 2017.
  - 20.- Arnaiz-Villena A, Palacio-Gruber J, **Muñiz E**, Rey D, Nikbin B, Nickman H, Campos C, Martín-Villa JM, Amirzargar A. *Origin of Azeris (Iran) according to HLA genes*. *International Journal of Modern Anthropology*. 10:115-138. 2017.
  - 21.- Arnaiz-Villena A, Palacio-Gruber J, Enriquez-de-Salamanca M, Juárez I, Campos C, Nieto J, **Muñiz E**, Martín-Villa JM. *HLA-G, -A haplotypes in Amerindians (Ecuador): HLA-G\*01:05N World distribution*. *Human immunology*. doi: 10.1016/j.humimm.2017.12.002. 2017.
  - 22.- Arnaiz-Villena A, Bayona B, Palacio-Gruber J, Hernández E, **Muñiz E**, Campos C, Juárez I, Gómez-Casado E, Martín-Villa JM, Silvera C. *HLA genes in Barranquilla (North Colombia): Searching for cryptic Amerindian genes*. *Human immunology*. 79:3-4. 2018
  - 23.- Arnaiz-Villena A, Enriquez-de-Salamanca M, Palacio-Gruber J, Juárez I, **Muñiz E**, Nieto J, Campos C, Martín-Villa JM. *HLA-G in Amerindians: Epidemiology and Worldwide Population Comparison*. *Open Medicine Journal*. 5:1-12.2018.
  - 24.- Arnaiz-Villena A, Palacio-Gruber J, Juárez I, Hernández E, **Muñiz E**, Bayona B, Campos C, Nieto J, Martín-Villa M, Silvera C. *HLA in North Colombia Chimila Amerindians*. *Human Immunology*. <https://doi.org/10.1016/j.humimm.2018.02.004>. 2018

### Comunicaciones a congresos

- 1- Arnaiz-Villena A, Serrano-Vela JI, Areces C, Rey D, Enriquez de Salamanca M, **Muñiz E**, Marco J, Gómez-Prieto P. *Songbirds conserved sites and intron size of MHC class I molecules reveal a unique evolution in vertebrates*. **XXII Congreso Español de Ornitología**. Madrid, España, 2014.
- 2- Arnaiz-Villena A, Enriquez-de-Salamanca M, **Muñiz E**, Basile L, Areces C, Carballo A, Martín-Villa M, Martínez-Quiles N, Rey D. *Easter Island and Aleuts relatedness to Amerindians and Pacific Islanders according to HLA genes*. **29<sup>th</sup> European Immunogenetics and Histocompatibility Conference**. Geneva, Switzerland. 2015.
- 3- Arnaiz-Villena A, Gómez-Prieto P, Enriquez de Salamanca M, Rey D, **Muñiz E**, Corrales A, Martín Villa M, Martínez-Quiles N, Areces C. *MHC variability in wild Songbirds: normal intron size and molecule evolution suggest that a minimal essential MHC does not exist in birds*. **29<sup>th</sup> European**

- Immunogenetics and Histocompatibility Conference.** Geneva, Switzerland. 2015.
- 4- Arnaiz-Villena A, Vargas-Alarcón G, Rey D, Enríquez-de-Salamanca M, **Muñiz E**, Areces C, Martín-Villa J, Martínez-Quiles N, Cruz-Robles D, López-Pacheco F. Mixtec Mexican Amerindians: an HLA alleles study for America peopling, pharmacogenomics and transplantation. **29<sup>th</sup> European Immunogenetics and Histocompatibility Conference.** Geneva, Switzerland. 2015.
  - 5- Arnaiz-Villena A, Ruiz-del-Valle V, Areces C, **Muñiz E**, Enriquez-de-Salamanca M. Songbirds conserved sites of MHC class I molecules reveal a unique evolution in vertebrates. **American Ornithologists' Union & Cooper Ornithological Society Conference.** Norman, Oklahoma. 2015.
  - 6- Arnaiz-Villena A, Ruiz-del-Valle V, Gomez-Prieto P, Areces C, Enriquez-de-Salamanca M, Basile L, Martín-Villa J, Martínez-Quiles N, **Muñiz E**. Carduelini new systematics: New groups and splits showed by mitochondrial DNA and Bayesian phylogeny. **American Ornithologists' Union & Cooper Ornithological Society Conference.** Norman, Oklahoma. 2015.
  - 7- Arnaiz-Villena A, Ruiz-del-Valle V, **Muñiz E**, Enriquez-de-Salamanca M, Gomez-Prieto P, Lowry E, Rey D, Basile L, Serrano-Vela JI, Areces C. Songbirds conserved sites of MHC Class I molecules reveal a unique evolution in vertebrates. **European Ornithologists' Union Conference.** Badajoz, España. 2015
  - 8- Arnaiz-Villena A, Ruiz-del-Valle V, **Muñiz E**, Enriquez-de-Salamanca M, Gomez-Prieto P, Lowry E, Rey D, Basile L, Serrano-Vela JI, Areces C. *Carduelini new systematics: New groups and splits showed by mitochondrial DNA and Bayesian phylogeny.***European Ornithologists' Union Conference.** Badajoz, España. 2015.
  - 9- Arnaiz-Villena A, Areces C, Enriquez-de-Salamanca M, Marco J, **Muñiz E**, Fernández-Honrado M, Rey D. *Evidence of Paleohistorical Atlantic and Pacific transoceanic genetic and cultural contacts.* **84<sup>th</sup> Annual Meeting of the American Association of Physical Anthropologists.** St. Louis. Missouri. 2015
  - 10- Arnaiz Villena A, Ruiz del Valle V, Gomez Prieto P, Palacio Grüber J, Campos C, Camacho A, Martin Villa M, Martinez Quiles N, Muñiz E. *Carduelini new systematics: New groups and splits showed by mitochondrial DNA and Bayesian phylogeny.* **III Iberian congress of Biological Systematics.** Vitoria-Gasteiz. España. 2015
  - 11- **Muñiz E**, Palacio-Grüber J, CamposC, Rey D, Tomasi S, Martín Villa JM, Arnaiz-Villena A. Easter Island and Aleuts relatedness to Amerindians and Pacific Islanders according to HLA genes. **39 Congreso de la Sociedad Española de Inmunología.** Alicante, España. 2016.
  - 12- Palacio-Grüber J, **Muñiz E**, Amirzargar A, CamposC, Rey D, Martín Villa JM, Arnaiz-Villena A. Kurds HLA genes: its implications in transplantation and pharmacogenomics. **39 Congreso de la Sociedad Española de Inmunología.** Alicante, España. 2016.



- 13- Arnaiz-Villena A, **Muñiz E**, Palacio-Grüber J, CamposC, Martín Villa JM, Gomez-Casado E. MHC variability in wild songbirds: normal intron size and molecule evolution suggest that a minimal essential MHC does not exist in birds. **39 Congreso de la Sociedad Española de Inmunología**. Alicante, España. 2016.
- 14- Vargas-Alarcón G, Cruz-Robles D, López-Pacheco F, **Muñiz E**, Palacio-Grüber J, CamposC, Rey D, Martín-Villa JM, Arnaiz-Villena A. Mixtec Mexican Amerindians: an HLA alleles study for America peopling, pharmacogenomics and transplantation.**39 Congreso de la Sociedad Española de Inmunología**. Alicante, España. 2016.
- 15- Arnaiz-Villena A, Palacio-Grüber J, Jose Recio M, Martinez-Laso J, Gomez-Casado E, Martin Villa JM, Martinez-Quiles N, Campos C, **Muñiz E**. Major Histocompatibility complex-DMB allelic diversity in old and new world nonhuman primates: Intraspecies pattern of evolution. **30th European Immunogenetics and Histocompatibility Conference**. Kos, Grecia. 2016.
- 16- Arnaiz-Villena A, **Muñiz E**, Campos C, Gomez-Casado E, Tomasi S, Martinez-Quiles N, Martin Villa M, Palacio-Grüber J. Origin of AncientCanaryIslanders Guanches: presence of Atlantic/Iberian Human Leukocyte Antigen and Y chromosome genes and Ancient Iberian language. **30th European Immunogenetics and Histocompatibility Conference**. Kos, Grecia. 2016.
- 17- Arnaiz-Villena A, Amirzargar A, **Muñiz E**, Campos C, Enríquez-de-Salamanca M, Rey D, Martinez-Quiles N, Martin-Villa JM, Palacio-Grüber J. Kurds Human LeukocyteAntigens Genes: ItsImplications in Transplantation and Pharmacogenomics. **30th European Immunogenetics and Histocompatibility Conference**. Kos, Grecia. 2016.
- 18- Arnaiz-Villena A, Muñiz E, Campos C, Martin-Villa JM, Palácio-Gruber J. Evidence of Prehistorical Atlantic and Pacific Transoceanic genetic and cultural contacts with America. **86<sup>th</sup> Annual Meeting of the American Association of Physical Anthropologists**. New Orleans, Louisiana. 2017.
- 19- Arnaiz-Villena A, Campos C, **Muñiz E**, Gomez-Casado E, Martín-Villa J.M, Palacio-Gruber J. First Population of Canary Island is not only North African: Presence of European Atlantic Genes and Ancient Iberian Language Scripts. **40 Congreso de la Sociedad Española de Inmunología**. Zaragoza, España. 2017.
- 20- Arnaiz-Villena A, **Muñiz E**, Palacio-Gruber J, Campos C, Alonso-Rubio J, Gomez-Casado E, Tejedor B, Martín-Villa JM, Silvera C. Pacific and European Ancestry of Amerindians: a HLA Relatedness Study in Wiwa (Arsario) Colombian Populations. **40 Congreso de la Sociedad Española de Inmunología**. Zaragoza, España. 2017.
- 21- Arnaiz-Villena A, Palacio-Gruber J, **Muñiz E**, Campos C, Alonso-Rubio J, Gomez-Casado E, Martín-Villa J.M, Silvera C. HLA Genes in Chimila Amerindians (Colombia), the Peopling of America and Relatedness to Europeans and Pacific Islanders. **40 Congreso de la Sociedad Española de Inmunología**. Zaragoza, España. 2017.

- 22- Arnaiz-Villena A, **Muñiz E**, Campos C, Palacio-Gruber J, Martín-Villa J.M, Tejedor B, Gomez-Casado E, Enriquez de Salamanca M. Characterization and Functional Implications of the Two New HLA-G Alleles Found in Amerindians (HLA-G\*01:08:02) and Caribbean (HLA-G\*01:20) Populations. **40 Congreso de la Sociedad Española de Inmunología**. Zaragoza, España. 2017.
- 23- Recio M.J, Palacio-Gruber J, Martinez-Laso J, **Muñiz E**, Gomez-Casado E, Martín-Villa J.M, Campos C, Arnaiz-Villena A. Major Histocompatibility Complex-DMB Allelic Diversity in Old and New World Nonhuman Primates: Intraspecies Pattern of Evolution. **40 Congreso de la Sociedad Española de Inmunología**. Zaragoza, España. 2017.
- 24- Arnaiz-Villena A, Palacio-Gruber J, **Muñiz E**, Recio M.J, Campos C, Tejedor B, Martín-Villa J.M, Martinez-Laso J. HLA-DMB in Amerindians: Specific Linkage of DMB\*01:03:01/DRB1 Alleles. **40 Congreso de la Sociedad Española de Inmunología**. Zaragoza, España. 2017.
- 25- Arnaiz-Villena A, **Muñiz E**, Campos C, Palacio-Gruber J, Martín-Villa J.M, Tejedor B, Gomez-Casado E, Enriquez de Salamanca M. Characterization and Functional Implications of the Two New HLA-G Alleles Found in Amerindians (HLA-G\*01:08:02) and Caribbean (HLA-G\*01:20) Populations. **31<sup>st</sup> European Immunogenetics and Histocompatibility Conference**. Mannheim, Germany. 2017.
- 26- Arnaiz-Villena A, Palacio-Gruber J, **Muñiz E**, Recio M.J, Campos C, Tejedor B, Martín-Villa J.M, Martinez-Laso J. HLA-DMB in Amerindians: Specific Linkage of DMB\*01:03:01/DRB1 Alleles. **31<sup>st</sup> European Immunogenetics and Histocompatibility Conference**. Mannheim, Germany. 2017.
- 27- Recio M.J, Palacio-Gruber J, Martinez-Laso J, **Muñiz E**, Gomez-Casado E, Martín-Villa J.M, Campos C, Arnaiz-Villena A. Major Histocompatibility Complex-DMB Allelic Diversity in Old and New World Nonhuman Primates: Intraspecies Pattern of Evolution. **31<sup>st</sup> European Immunogenetics and Histocompatibility Conference**. Mannheim, Germany. 2017.
- 28- Arnaiz-Villena A, Campos C, **Muñiz E**, Gomez-Casado E, Martín-Villa J.M, Palacio-Gruber J. First Population of Canary Island is not only North African: Presence of European Atlantic Genes and Ancient Iberian Language Scripts. **31<sup>st</sup> European Immunogenetics and Histocompatibility Conference**. Mannheim, Germany. 2017.
- 29- Arnaiz-Villena A, **Muñiz E**, Palacio-Gruber J, Campos C, Alonso-Rubio J, Gomez-Casado E, Tejedor B, Martín-Villa J.M, Silvera C. Pacific and European Ancestry of Amerindians: a HLA Relatedness Study in Wiwa (Arsario) Colombian Populations. **31<sup>st</sup> European Immunogenetics and Histocompatibility Conference**. Mannheim, Germany. 2017.

- 30- Arnaiz-Villena A, Palacio-Gruber J, **Muñiz E**, Campos C, Alonso-Rubio J, Gomez-Casado E, Martín-Villa J.M, Silvera C. HLA Genes in Chimila Amerindians (Colombia), the Peopling of America and Relatedness to Europeans and Pacific Islanders. **31<sup>st</sup> European Immunogenetics and Histocompatibility Conference**. Mannheim, Germany. 2017.